CRWR Online Report 97-2

Preliminary Investigation of Implementing an Environmental Information System for the Pantex Facility

by

Ye Maggie Ruan, MA

Graduate Research Assistant

and

David R. Maidment, PhD.

Principal Investigator

December, 1997

CENTER FOR RESEARCH IN WATER RESOURCES

Bureau of Engineering Research • The University of Texas at Austin

J.J. Pickle Research Campus • Austin, TX 78712-4497

This document is available online via World Wide Web at

http://www.ce.utexas.edu/centers/crwr/reports/online.html

Acknowledgements

1997 is an unforgettable year for me. What do I really want to do as my career? I have been asked and also have asked myself this question many times. I have found the answer in 1997 and in Dr. Maidment's research group.

I do feel lucky and proud to be part of this group. In this group, we not only help each other, we also build each other. I have learned so much from every one in the group this year, I want to say: Thank You, Y' All!

I want especially to thank Seann Reed, who is always a great help in many aspects from simple Avenue scripts to complicated methodologies. Thank you, Seann, this year would have been different without your help.

I want to also thank Tim Plover, our system administrator. Tim, I do think you have magic fingers. They not only fix all system problems, but also open the exciting computer world to me. Thank you, Tim.

And most importantly, I want to thank my supervisor Dr. Maidment for his guidance and encouragement. What I learned from him is not only the research approach, but also to be a positive and enthusiastic human being. Dr. Maidment, I do think you bring the best out in me. Now, everything is "doable" and will be in "good shape". Thank you, Dr. Maidment.

Thanks also go to my co-advisor Dr. Randall Charbeneau and Amarillo National Resource Center for Plutonium who provided the funding for this project.

Abstract

The Pantex facility is America's only nuclear weapon assembly and disassembly site. Environmental and safety issues are of critical importance for the Pantex facility and the surrounding region. An Environmental Information System framework has been proposed to help to study these issues.

In this report, the concept of the Environmental Information System framework is introduced. The essential components of the Environmental Information System are described in detail. The key component in this framework is a spatial database, which describes the environment of the Pantex facility and the surrounding region.

Preliminary investigations on some of the components of a Pantex Environmental Information System are presented. These investigations include 1) developing a spatial database that covers the Pantex facility and the surrounding region and archiving the database in CD-ROM format, 2) integrating MS Access as an external relational database to ArcView, 3) investigating the feasibility of implementing an Internet Map Server, and 4) setting up an research intranet for Risk Assessment of Pantex Plutonium MOX Fuel Facility.

The information presented by a Pantex Environmental Information System can assist the environmental and operational projects of the Pantex facility, and also be beneficial for the general public to understand the environmental issues related to Pantex.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS III
ABSTRACT IV
LIST OF TABLES VII
LIST OF FIGURES VIII
CHAPTER 1: INTRODUCTION 1
1.1 Background 1
1.1.1 Pantex Background 1
1.1.2 Environmental Issues 2
1.1.3 Environmental Monitoring 5
1.1.4 Public Concern 6
1.2 Objectives and Scope 6
1.2.1 Objectives 6
1.2.2 Scope 7
1.3 Research Approach 9
CHAPTER 2: LITERATURE REVIEW 11
2.1 Current Status of Pantex Environmental Information Database 11
2.2 Environmental database development and management 13
2.3 Environmental Information Distribution 15
2.4 Summary 16
CHAPTER 3: METHODOLOGY 18
3.1 Definition of Environmental Information System 18
3.2 Framework of Environmental Information System 19
3.2.1 Spatial Database Development 20
3.2.1.1 Information Sources 20
3.2.1.2 Spatial Database Design 25
3.2.2 Spatial Database Management 34
3.2.2.1 Spatial Database Administration 34
3.2.2.1 Spatial Database Distribution 38
3.2.3 Spatial Database Implementation and Training 44
3.2.3.1 Implementation 44
3.2.3.2 Training 46
CHAPTER 4: SPATIAL DATABASE DEVELOPMENT FOR PANTEX ENVIRONMENTAL INFORMATION SYSTEM 48
4.1 Raster Image Data 52
4.1.1 Retrieving the Image from Sure!Map Raster 52
4.1.2 Converting the Image to Grid and Projecting to State Plane Coordinates 53
4.2 Census Data 54
4.2.1 Data Acquisition 55
4.2.2 Data Processing 57

TABLE OF CONTENTS (CONT.)

4.3 Stream and Transportation Data 67
4.4 Land Use Data 73
4.4.1Retrieving the Source Data 73
4.4.2 Merging the Data Sets 74
4.4.3 Correcting Mapping Errors using ArcEdit 75
4.5 Soils Data 78
4.6 Vegetation Data 80
4.6.1 USGS LUCC 80
4.6.2 Vegetation Data from TNRIS 85
4.7 Aquifer Data 87
4.8 Pantex Facility Data (CAD) 88
CHAPTER 5: ADDITIONAL DATABASE COMPONENTS AND TECHNOLOGIES 94
5.1 Integrating Microsoft Access with ArcView 94
5.1.1 Introduction 94
5.1.2 Objectives 96
5.1.3 Methodology 96
5.1.4 Procedure and Results 97
5.2 Internet Map Server 100
5.2.1 MapObjects Internet Map Server 100
5.2.2 ArcView Internet Map Server 102
5.2.3 Comparison of the MapObjects and ArcView Internet Map Servers 105
5.3 Pantex MOX Fuel Facility Risk Assessment Research Intranet 107
CHAPTER 6: CONCLUSION 110
REFERENCES 113
Appendix A Projection Files 116
Appendix B Aml Script 124
Appendix C Avenue Scripts 128
Appendix D Olson Global Ecosystems Legend 133
Appendix E Pantex Spatial Database CD-ROM Documentation 137
Appendix F Integrating Microsoft Access to ArcView 155

List of Tables

Table 4.1 Names of the Twenty Counties in the Study Region and The Fips Codes 51

Table 4.2 Composition of BLCKGR-ID in TIGER Data and Texas Census Table. 58

Table 5.1 Differences between ArcView IMS and MapObjects IMS 106

Table A-1 The Projection Files Used in the Pantex Spatial Database Development 117

List of Figures

Figure 3.1. Environmental Information System Framework 20

Figure 4.1 Technical Subjects of Risk Characterization of Pantex MOX Fuel Facility...48

Figure 4.2 The Study Region and the Projection of the Spatial Database 51

Figure 4.3 Census Data in Block Level for Twenty Counties..............................64

Figure 4.4 Census Data at the Block Level for the Study Region........................65

Figure 4.4.a The Population of Children in the Study Region with an Age between Zero to Four..66

Figure 4.4.b The Age Distribution in the Study Region 66

Figure 4.5 Hydrography (DLG) of the Study Region...71

Figure 4.6 Road (DLG) of the Study Region 71

Figure 4.7 Railroad (DLG) of the Study Region 72

Figure 4.8 Miscellaneous Transportation (e.g. Airport and pipelines) 72

Figure 4.9 Land Use Data in the Study Region..78

Figure 4.10 STATSGO Soil Data for the Study Region.....................................79

Figure 4.11 Vegetation Description of the Study Region Provided by GLCC 85

Figure 4.13 Aquifer Data in the Study Region...87

Figure 4.14 Pantex Building CAD Map in ArcView..89

Figure 4.15 Pantex 5-ft Contour Map 89

Figure 4.16 Pantex 1-ft Contour Map 90

Figure 4.17 Pantex USGS Quadrangle Map...90

Figure 4.18 Pantex Plant Outline (Low Detail) 91

Figure 4.19 Eight Data Layers in the Pantex Spatial Database 92

Figure 5.1 Schematic Diagram of the Integration of Access with ArcView.........96

Figure 5.2 Nitrate Concentration Contours in the Groundwater of the Pantex Facility..99

Figure 5.3 MapObjects Internet Map Server (ESRI White Paper Series, 1997b) 101

Figure 5.4 ArcView Internet Map Server..103

Figure 5.5 Sample web page created by ArcView IMS (ESRI White Paper Series, 1997c) 104

Chapter 1: INTRODUCTION

1.1 BACKGROUND

The Pantex facility is America's only nuclear weapon assembly and disassembly site. Environmental and safety issues are of critical importance for the Pantex facility and the surrounding region. The Environmental Impact Statement of the Pantex facility has a list of environmental issues to be examined. To examine these issues, various types of environmental information are required. Since the use and the disposal of nuclear weapons and radioactive materials are very sensitive issues to the public, a smooth and clear communication between the Pantex facility and the public is also essential. An Environmental Information System framework has been proposed to help address the environmental issues. The key component in this framework is a spatial database. The spatial database is the environmental description of the Pantex facility and the surrounding region. The information presented by the spatial database will not only assist the environmental and operational projects of the Pantex facility, but also be beneficial for the general public to understand the environmental issues that related to the Pantex.

1.1.1 Pantex Background

Currently, there is only one nuclear weapons assembly and disassembly facility in the United States, the Pantex Plant. The facility is located on the High

Plains of the Texas Panhandle, 17miles northeast of Amarillo, centered on a 16,000-acre site just north of U. S. Highway 60 in Carson County.

The Pantex Plant was originally constructed as a conventional bomb plant for the U.S. Army during the early days of World War II. After the war ended, Pantex was deactivated. In 1951, at the request of the Atomic Energy Commission (now the Department of Energy), the Army reclaimed the main plant and 10,000 surrounding acres for use as a nuclear weapons production facility.

During the 1960's, several other nuclear weapon operations were closed, leaving Pantex the only nuclear weapons assembly and disassembly facility in the US. In the 1990's, as one of the results of the end of the cold war, the United States and the former Soviet Union are now working to reduce their nuclear weapons stockpiles, and Pantex plays a vital part in this operation. The Department of Energy (DOE) has assigned all current dismantlement activities to the Pantex Plant, which means that weapons will be dismantled on a schedule of up to 2,000 per year until the stockpile has decreased to the predetermined number. It is estimated that dismantlement should be completed by 2004. (Pantex, 1997a)

1.1.2 Environmental Issues

According to the Pantex web site, (Pantex, 1997b) disassembly and disposition operations at Pantex are conducted under the highest possible levels of safety and security. Protecting the environment and safeguarding human safety and health are of paramount importance to the people at Pantex and the surrounding area. The management of the Pantex Plant is committed to a policy of openness regarding these issues. Additional information about Pantex and other issues related to radiation, environmental impact, environmental safety and health, and interim storage can be retrieved from this web site.

The Environmental Impact Statement is a document which describes the impacts on the environment as a result of a proposed action as well as plans to mitigate these impacts. In the Pantex Environmental Impact Statement (Pantex web site, c), a few of major environmental issues are listed below:

- The usage of Pantex as a possible site for storing nuclear components that result from weapons disassembly.
- Potential effects on the public and workers from releases of radiological and hazardous materials during normal operations and from accidents, including aircraft crashes.
- Effects of natural disasters including floods, tornadoes, and earth- quakes.
- Potential effect on air and water quality and other environmental consequences of normal operations and potential accidents--including any effects on the Ogallala Aquifer.
- Potential effects of operation at the Pantex Plant, including relevant impacts from transportation activities, environmental restoration, and present operations.

- Potential effects on endangered species, economically and recreationally important species, floodplains, wetlands, and historic and archaeological resources including paleontological sites and Native American resources.
- Potential effects on future decontamination and decommissioning decisions.
- Potential socioeconomic impacts on communities near the Pantex Plant, including demographics, economic base, labor pool, housing, transportation, utilities, public services/facilities, education, and environmental justice.
- Effects on near and long-term waste management practices and activities including pollution prevention and waste minimization.
- Potential effects on agricultural lands and practices.
- Potential impacts of noise levels on the surrounding environment and the people living there.
- Potential impacts on scenic and visual resources.
- Potential impacts on land use plans, policies, and controls.

In order to study these issues, many universities, research institutes and environmental consulting firms have conducted environmental projects. Those projects usually require various types of environmental information, such as land description, census information, and environmental monitoring data. The land description may include data types, such as land use, soil types, vegetation, elevation, stream, lake and aquifer distribution.

1.1.3 Environmental Monitoring

The Pantex Plant has conducted environmental monitoring for over 20 years. The present monitoring program includes two principal collection and analysis activities:

- Monitoring liquid and airborne discharges to describe and measure releases
- Environmental surveillance of air, water, soil, and vegetation, and measurement of external radiation to evaluate the environmental impacts of Pantex Plant operations

There are a total of around 6000 data sampling locations for air, ground water, surface water, soil, vegetation, and animal monitoring. All of the monitoring programs are validated by quality assurance programs. Currently there is about 2.4 gigabytes of environmental sampling data accumulated at Pantex.

Since collecting and compiling the environmental information is a very time consuming process, many environmental projects would like to use the information that is readily available. These projects act as environmental information consumers. On the other hand, some other projects may actually generate new environmental information or update the originally available information. These projects act as environmental information suppliers. It would be beneficial for both parties to set up a data center in the form of an environmental spatial database, so that the environmental information suppliers and consumers could share information resources. Thus, the information resource from the supplier would not be wasted or limited by the distribution range, while the information consumer can access the data in a timely manner.

1.1.4 Public Concern

Nuclear weapons and radioactive materials are very sensitive issues for the public and the local government. The Pantex Plant is a unique facility with complex and potentially dangerous missions. In addition, it is close to a large city Amarillo with 206,878 population (Amarillo Demographics, 1997) and is located on top of a major aquifer that is the water supply for Amarillo. Communications and education between the Pantex Plant and general publics, community leaders, Pantex employees, local and Texas news media, state officers, and others are extremely important. This situation suggests setting up a smooth communication channel between the Pantex Plant and the public: where environmental information and education material can be easily accessible for the public, and the Pantex Plant can easily hear public feedback.

1.2 OBJECTIVES AND SCOPE

1.2.1 Objectives

The objective of the project is to conduct a preliminary investigation of implementing an Environmental Information System (EIS) framework for the Pantex Plant. The major component of the framework is a spatial database, which contains various environmental data that are related to the Pantex Plant and the surrounding region. The spatial database should be developed and managed in the format that it is accurate, updated, and easily accessible by any environmental project managers or researchers. The spatial database should also provide the public with the general environmental information and the necessary educational materials.

1.2.2 Scope

The scope of the project involves the following components: the importance of an Environmental Information Systems to the Pantex Plant, its relation to the environmental spatial database, the essential components of an Environment Information System, the function of each component, and the framework of implementing an Environmental Information System into the Pantex Plant.

The Environmental Information System framework, based on Geographic Information Systems (GIS), consists of two major components, the spatial database development and the spatial database management. The related issues in the spatial database development are the information sources, retrieving and collecting methods, data process procedures, and quality assurance. While in the spatial database management, the related issues are database administration and distribution. Other related issues are the hardware and software requirements of implementation of the framework and the training and technical support for the framework.

In this report, the concept of the Environmental Information System will be defined, and the design of the Environmental Information System framework will be outlined. Each of the components of the framework will be introduced, and its functionality will be clarified. A great amount of work has been done to implement the Environmental Information System framework. This includes the following projects:

- Building a Regional Spatial database for the Pantex Plant and the Surrounding Region
- Exploring a CD-ROM Approach of Spatial Database Development and Management using Microsoft Access and ArcView.
- Evaluation of Internet Map Server as Spatial Database Distribution Alternative.
- Setting up a Research Intranet for Pantex Plutonium Mixed-Oxide Fuel Process Facility

These projects serve as the examples of the individual components of the Environmental Information System framework for the Pantex Plant.

The current spatial database contains only the regional description data. Future works include implementing the environmental sampling data into the spatial database using Microsoft Access as an external relational database, providing more types of regional description data, setting up the Internet Map Server and refining the Research Intranet.

The final product of this report is presented in CD-ROM format. The CD-ROM includes: 1) the documents presented in Microsoft Word, Adobe Acrobat, and HTML formats, 2) the preliminary version of the actual spatial database and its documentation, 3) related projects for integrating MS Access to ArcView, 4) the original files and file structure of research Intranet. Most of the data sources used for the spatial database are linked through HTML documents.

1.3 RESEARCH APPROACH

The initial objective of project was to build the spatial database which would serve as a data source for Risk Assessment of the Pantex Mixed-Oxide Fuel Facility. In the process of developing the spatial database, we noted the importance of developing a framework to support the usage of the spatial database; otherwise we would not be able to get the most benefit out of the database. Since the purpose of building a spatial database is primarily for the environmental issues, we defined the framework as an Environmental Information System framework.

A literature review was conducted to see if similar types of frameworks have been developed. In addition, the information of the current environmental database at Pantex was gathered from the Pantex GIS Specialist Gary Thomas and Database Administrator Rob Fitch.

A preliminary spatial database was developed, which contains various types of regional spatial data and the Pantex facility spatial distribution data. Most data were processed and displayed using ARC/INFO and ArcView, the major GIS software provided by Environmental System Research Institute (ESRI). A new method of managing the GIS database using Microsoft Access was also investigated. The feasibility of setting up an Internet Map Server was evaluated. A Research Internet was set up to serve as one of the alternatives to distribute the environmental information.

In the following chapter, a brief overview of the current status of the Pantex environmental database and the literature review are provided. In Chapter 3, the concept of the Environmental Information System and its components are introduced. In Chapter 4, the detailed processes of constructing the Pantex Spatial Database and the final results are presented. In Chapter 5, the additional projects that relate to the Pantex Environmental Information System are presented. Finally the conclusions and the future works are presented in Chapter 6.

Chapter 2: LITERATURE REVIEW

In this chapter, the overview of the current status of Pantex Environmental Database is provided. Several examples of environmental information systems and their benefits are introduced. A summery of the literature review and the unique features of a CD-ROM approach are presented.

2.1 CURRENT STATUS OF PANTEX ENVIRONMENTAL INFORMATION DATABASE

The Pantex facility currently has an environmental database to serve as the information center of the environmental and operational projects. The information about the current status of the Pantex environmental database has been gathered from the Pantex GIS Specialist Gary Thomas and Database Administrator Rob Fitch (Thomas, Gary and Fitch, Rob).

The current database contains information about environmental samples, regulatory limits, numerous lookup (or validation tables), analytical labs, test costs, lab invoices, aquifers, bore holes, wells, well construction, conductivity logs, geological faults and penetrations, gamma logs, land use, log curves, maps, plumes, contaminant boundaries, geology, and hydrology. The maps themselves are not part of the database, but information about the maps is contained in the database.

The database management system used is Sybase version 11. Data are typically transferred using ASCII delimited files (to avoid compatibility problems). Data can be also directly transferred using Microsoft Access data transfer tools. The GIS software used at Pantex is the Intergraph MGE suite.

The database has 197 tables (see breakdown below) and 3,566,104 data records. It requires 2.4GB of disk storage. Database design started in February of 1994.

The components of the Pantex Environmental Database:

- 83 tables for environmental data
- 60 tables for the MGE interface
- 5 tables for other interfaces
- 28 special purpose tables
- 21 Sybase internal tables

The Pantex environmental database currently has information on 6,657 sampling locations. The database administrators track any sample taken at Pantex or around Pantex regardless of who takes the sample. Some samples are taken by Battelle Memorial, Argonne National Lab, U.S. Army Corps of Engineers, and TNRCC.

The database is updated daily. The database administrators receives electronic data from Pantex contracted analysis labs. Using locally developed software, the data is checked for validation errors and then loaded into temporary tables. Each record is then individually verified and validated online by the owners of the data. Once validated, the record is added to permanent tables and is available for retrieval or display by other users.

The database is currently used for operational, environmental, and research projects by about 30 local users and four external organizations. The local users are primarily from three different departments at Pantex (Environmental Restoration, Environmental Protection, and Waste Management).

Online access to the database is limited to the Pantex network. Any Pantex employee can get online access to the database by using any one of 14 different software products/tools. Because of the Pantex security system, external users are not given online access to the database. However, the Pantex database administrators can provide data dumps to any approved external users. The data dumps can be put on CD-ROM if requested.

2.2 ENVIRONMENTAL DATABASE DEVELOPMENT AND MANAGEMENT

The environmental database is complex due to the large amounts of the data input and analysis on a regular basis. It is very dynamic and needs to be continually updated. Therefore, to develop and manage an environmental database is a very challenging task.

Since most environmental information is spatially distributed, managing and analyzing environmental information in GIS may be an efficient choice. Many institutes, organizations and government agencies have been working on implementing the environmental information system for their researches and daily operations. The following are two examples.

PRC Environmental Management, Inc. (PRC) from Denver (http://www.ttemi.com) is using Oracle, ARC/INFO, ArcCAD, and ArcView software to support environmental restoration activities including site characterization, fate and transport modeling, human health and ecological risk assessment, remedial design, and long-term monitoring. The complex and multi-dimensional nature of data from potentially contaminated sites presents significant challenges for designers of environmental databases. PRC developed a system for maintaining attribute data in a normalized data structure and ensuring data integrity through the encapsulation of stored triggers, procedures, methods, constraints, and views. The triggers act as methods in the standard sense of the object-oriented programming model. When an insert, update, or delete message is sent to a database table (object) that is encapsulated with a constraint or trigger, the constraint is activated or the trigger executes to verify that the requested action does not violate data integrity parameters. Because environmental restoration programs frequently produce large volumes of chemical, geological, and geotechnical data, sophisticated performance tuning measures are used by PRC to retrieve the needed data in a timely manner. Through ESRI's database integrator, the data which are optimally stored in Oracle can be fully integrated with spatial data such as sample locations, installation restoration boundaries, parcels, land use, habitat delineations, and digital aerial photographs which are maintained in ARC/INFO. (ESRI User conference proceeding, 1997a)

In Hungary, the Ministry of Environment and Regional Policy implemented an information system integrating environmental and nature conservation data with GIS capability at the Information Center of the Ministry, as well as at twenty-two regional offices throughout the country. ESRI's ARC/INFO was selected as the high-end GIS with ArcView for data management, query, and analysis. Oracle Relational Database Management System is the tool for tabular data management. The concept was to develop a system facilitating collection of data, as well as processing, analyzing and decision making activities of the environmental and nature conservation sectors of the Ministry and supporting the simultaneous handling of data coming from the two sectors. The basic aspect of system development was to integrate a large amount of databases (e.g., emission, hazardous wastes, groundwater, sewage water, noise, biological observation data, cadastral data) into a uniform system with powerful GIS capability. (BOZÓ *et al*, 1997)

Many environmental projects make use of an environmental database to perform environmental impact assessment, environmental monitoring and modeling, environmental planing and management. The most recent development and application can be viewed in the ESRI user conference proceeding. (ESRI user conference proceeding, 1997)

2.3 Environmental Information Distribution

Environmental information should not only be available to the professional environmental engineers and scientists, but also to the general public. One of the examples of distributing environmental information to the public is ENVIROCITY – a Public Online Environmental Information Service for

European Cities. ENVIROCITY delivers environmental information to the urban public by means of telematics. The system is based on the existing GIS and telematics technologies. The project delivers a user requirement analysis, a GIS application in the form of a pre-prototype. The demonstrator will be developed on air quality data for the City of Munich and will further be implemented by the cities of Antwerp, Victoria, Piraeus and Lamia. (Schaller et al, 1996)

2.4 SUMMARY

The following conclusions were drawn from the literature review:

1. An Environmental database is beneficial for many environmental projects such as site characterization, fate and transport modeling, human health and ecological risk assessment, remedial design, and long-term monitoring. A well-organized environmental database should be set up to connect the data contributors and data requesters.
2. Environmental data are complex with multiple dimensions of time, media, phase, and geographic location, including depth. Databases in excess of one million analytical results and over 10 million data items are common.
3. Most environmental databases use large relational database software (e.g., Oracle, Sybase) to maintain the tabular sampling data, with the aid of GIS software (e.g., ARC/INFO, ArcView, MGE) to perform spatial analysis.

4. Environmental information should be provided not only to professional environmental engineers and scientists, but also to the general public.
5. The Pantex environmental database has a complicated data structure with 197 data tables, which contain information from 6,657 sampling locations. The database provides environmental information to various users for operational, environmental, and research projects of different organizations.

The summary of the literature review indicates that Environmental Information System of Pantex should be sufficient enough to support the complicated database structure and the wide range of users. In this report, a CD-ROM approach as a supplementary component of Environmental Information System is proposed. This CD-ROM approach can be a distribution alternative for the server end of the database, and the management alternative for the client end of the database. Using this approach, the database integrity of the server can be ensured, and the distribution range of the database can be increased. A customized, project specific spatial database with a simpler structure is possible for the users. The GIS functionality can be performed individually, therefore the dependence on the server and the network performance can be diminished.

Chapter 3: METHODOLOGY

In this chapter, the concept of the Environmental Information System is defined. The framework of the Environmental Information System is outlined and each of the components in the Environmental Information System is introduced.

3.1 DEFINITION OF ENVIRONMENTAL INFORMATION SYSTEM

The environment has become a critical issue worldwide as we enter the twenty-first century. This century is also being called “The Information Age” since its progress and politics have been guided by control of information in the same way that development during the “Industrial Age” was guided by control over industry and mass production. An Environmental Information System is defined as a computer-based system for the storage, management, and analysis of environmental information and associated data. An Environmental Information System contains information such as land surface description (e.g., streams, transportation links, soil, land use, vegetation, geological faults and penetrations, etc.), subsurface description (e.g., groundwater, mining, etc.), environmental operation records (e.g., bore holes, wells, well construction, conductivity logs, etc.), environmental monitoring records (e.g. environmental sampling data, plumes, contaminant boundaries, etc.), regulatory limits, geology, hydrology (e.g. precipitation, evaporation, temperature, radiation, wind speed), and documents and descriptions of related projects (e.g., Environmental Impact Statement, maps, etc.).

The core component of an Environmental Information System is a well structured and easy-accessible spatial database, which contains spatially distributed information with its related attribute information. The purpose of an Environmental Information System is to provide the necessary environmental information and regulatory criteria to the environmental project managers or researchers, regulatory agencies, and local or federal government. The Environmental Information System could also serve as an information center to the general public for the environmental awareness. The Environmental Information Systems are constructed, maintained, and distributed through various environmental information technologies.

3.2 FRAMEWORK OF ENVIRONMENTAL INFORMATION SYSTEM

Since the core component of an environmental information system is a spatial database, the main tasks of implementing an Environmental Information System are the spatial database development and management. The framework should also include the implementation and training technologies. The framework of the Environmental Information System is described in the Figure 3.1.

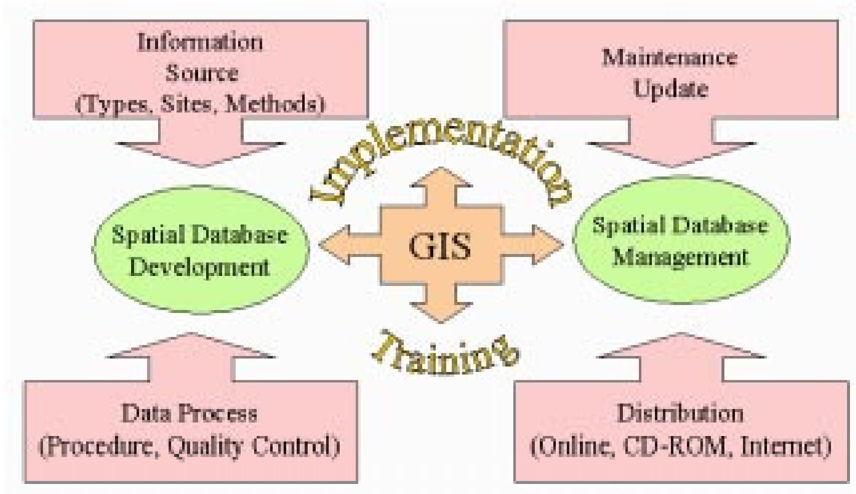


Figure 3.1. Environmental Information System Framework

3.2.1 Spatial Database Development

Spatial database development includes two major components: gathering environmental information from various resources and processing the information or data sets into a consistent format so that they can be easily accessed and utilized.

3.2.1.1 Information Sources

Types of environmental information

Environmental data can be divided into three major categories: regional description data, environmental operation and sampling data and regulatory limit data.

Regional description data describe the regional geographic features and their related attribute information. Geographic features describe the objects or phenomena commonly seen in the natural and man-made environment (e.g., roads, streams, lakes, elevation, land use, soil, vegetation, aquifer distribution, site layout, etc.). Regional description data are usually collected by large surveys conducted by government agencies and data venders. These data should be collected using standard methods and processed into a consistent format over a relatively large region. The data should also be examined by a stringent quality control process before they are provided to the public. Since there is a tremendous effort spent in one large survey, regional geographic data are less likely to have time series associated with them. However, some regional geographic features (e.g., roads, buildings, land use, etc.) do change over the time, so it is necessary to get the most current data in order to obtain reliable results.

Environmental operation and sampling data refer to those data that are collected on a regular basis from scattered operational or sampling locations, as well as the description of those location and data (e.g., precipitation, runoff, evaporation, environmental sampling data, plumes, contaminant boundaries, boreholes, wells, well construction, conductivity logs, etc.). These data are usually collected by government agencies (e.g., USGS gauge stations for stream flow), research institutes (e.g., Oregon State University for precipitation data set, PRISM) and individual facilities (e.g., Pantex for environmental sampling data). These data are entered into the databases regularly, which requires open, easily updated databases.

Regulatory limit data are the criteria set by government agencies (e.g. EPA) who conduct extensive studies on environmental impact and environmental health and safety. These criteria represent constituent concentrations or levels associated with a degree of environmental effect upon which scientific judgement may be based. Definitions of acceptable quality may relate to a unique local situation involving political, economic and social factors, but most of them apply more generally. These data are static, although they may change under specific circumstance, such as with a change in environmental laws.

Sources of environmental information

Most of the regional geographic data are provided by government agencies. The data are usually provided in the format of CD-ROM, which can be ordered. With the maturing of computer technology both in high-speed personal computers and the "information superhighway"- the Internet, many data can be readily downloaded directly from the Internet. These sites are usually maintained by the following organizations:

- Government agencies
- International agencies
- Chambers of Commerce
- Local libraries
- Private vendors
- Other GIS users

Here are some of the sites that are useful for searching for regional description data sets:

USGS-US GeoData (http://edcwww.cr.usgs.gov/doc/edchome/ndcdb/ndcdb.html)

- 1:250,000-Scale Digital Elevation Model (DEM)
- 1:2,000,000-Scale Digital Line Graphs (DLG) - SDTS format only
- 1:100,000-Scale Digital Line Graphs (DLG)
- Large Scale Digital Line Graphs (DLG) - SDTS format Only
- 1:250,000-Scale and 1:100,000-Scale Land Use and Land Cover (LULC)

Texas Natural Resource Information System (TNRIS) Anonymous FTP site (http://www.tnris.state.tx.us/ftparea.html)

This site provides a great amount of data sets in Texas for various purposes, as well as some utility software for the use of the data sets.

The USEPA Reach Files (http://www.epa.gov/OWOW/NPS/rf/)

The U.S. Environmental Protection Agency has generated a new version of the Reach File Version 1.0 (RF1) and metadata for this national hydrologic data set in ARCINFO format. You can link to the new RF1 metadata and the USEPA Reach Files homepage: USEPA-NSDI node for RF1 metadata (http://nsdi.epa.gov/nsdi/projects/rf1_meta.html).

USGS-National Geospatial Data Clearinghouse (http://nsdi.usgs.gov/nsdi/products/huc.html)

This site provides Hydrologic Unit Maps in 1:2,000,000 and 1:250,000 scale.

National Climatic Data Center (http://www.ncdc.noaa.gov/)

Dr. David Maidment, professor of University of Texas at Austin, has many useful Internet sites linked in his web site (http://www.ce.utexas.edu/prof/maidment).

Other regional geographic data may be provided in CD-ROM format by data vendors who gather the data by scanning, digitizing, and aerial photography. The data types are usually raster image map and digital orthophotograph. The raster image map used in the Pantex spatial database was purchased from the Horizons Inc. company (http://www.horizons.com).

For more GIS related web site, please check GIS Resource page at http://www.ce.utexas.edu/prof/maidment/gisrsrc.html

The Environmental System Research Institute (ESRI) also provides an extensive set of ready-to-use data, ESRI Data and Maps, Volume 1 CD-ROM, accompanying several ESRI software products. The data are delivered in shape file (.shp) and other popular formats for immediate use with ESRI software. The CD-ROM provides geographic features (e.g. lakes, rivers, road, political boundaries, census tracts, zip codes, etc.) of the US, Canada, Mexico and the World.

Many of the local data can also be generated in house. For example, the site facility CAD file may be converted to the format that can be directly readable by ArcView. Site maps can be scanned or digitized to be stored in a spatial database.

In the research, a preliminary spatial database has been developed for the Pantex facility and the surrounding area, for the purpose of Risk Characterization of the Pantex Plutonium Mixed-Oxide Fuel facility. According the characteristics of this project, the following data are included in the spatial database:

- Aquifer description
- Stream description
- Transportation description
- Land use description
- Soil description
- Vegetation description
- Census description of the population
- Facility layout

The Pantex Spatial Database is provided in CD-ROM format. The documentation of the CD-ROM can be reached at the web site: http://www.ce.utexas.edu/prof/maidment/intranet/pantex/pubwin/CD-ROMdoc.html.

3.2.1.2 Spatial Database Design

A well-organized spatial database should contain data in a consistent format for a specifically defined geographic region, however, the information retrieved from various sources are usually in different formats and cover different regions. This raw information needs to be processed using standard conversion methods and procedures. The final data and the process itself should be examined and monitored for the purpose of quality control. It is recommended to provide a metadata document along with the data that will be stored in a spatial database.

Design Procedures

A well-designed spatial database should provide a comprehensive theoretical framework and organization. Users should be involved in the design process. The database should allow users to view the database in its entirety and evaluate how the various aspects of the database need to interact. This forces users to identify major issues, potential problems and organizational constraints. It should satisfy users' needs and be application-independent. It should also be consistent and easy to use.

A typical design procedure includes determining the database contents, selecting appropriate geographic data sets, organizing the contents into a series of themes, and documenting the design procedure and final results. Normally the design procedure can be separated into two phases, conceptual design and physical design.

Phase I Conceptual Design

The elements in the conceptual design phase are assessing the users' needs, researching and evaluating the data resources and defining the study region and projection type.

Assess the Users' Needs

The database should be structured to meet the users' needs. Data should be organized based on the inherent characteristics of the data, instead of by specific uses or applications. The data structure must support all users' needs, based on the general set of applications identified in the needs assessment. Based on our experience, the following questions should be answered by the users' need assessment:

1. Who are the users?
2. Where are the users located?
3. What types of the data are required?
4. What are the resolution, precision and accuracy requirements of the data?
5. How will the users retrieve the information from the database?
6. How often will the users retrieve the data?

Research the Data Resources

The data types and the data resolution, precision and accuracy requirements need to be defined before starting the research on the data resources. The following options can be used for researching the data resources:

1. Examine the inventory of in-house data sources
2. Generate your own data by scanning, global position system (GPS), COGO and digitizing, and aerial photography.
3. Explore the Internet
4. Purchase from data vendors

Evaluate the Data Sources

The following considerations should be evaluated for each data source:

1. Map projection - All of the data sets have to be projected from their original projection to a consistent map projection defined for the spatial database
2. Scale – The amount of detail necessary for the GIS application is determined by the scale of the source data which, in turn, affects the type of attributes available. Scale is the measure of the resolution of the GIS data.
3. Accuracy – Some applications demand greater locational accuracy than others.
4. Compatibility – The format of data you buy or create must be compatible other data and with the GIS software the users employ. The compatibility of the data is largely determined by the adoption and implementation of data standards. Data standards bring order to the data development process and should be agreed upon by all GIS program participants. There are many data standards in the GIS industry. The most recently discussed data standard is Spatial Data Transfer Standard (SDTS). (ESRI Educational Services, 1996a)
5. Precision – What is the precision of digital data sources? Does it meet your application needs?
6. Resolution – The smallest object that can be represented at a given scale (e.g., objects less than 2 acres are generally depicted as points on 1:24,000 scale maps)

7. Age of the data – Some geographic features do change over time, such as new streets, buildings, land use distributions and vegetation distributions. Current and updated information is necessary for a reliable database.

Define the study region

The study region should be defined based on user needs before any of the following processes. There are two main purposes for defining a study region. First, the scale and resolution of the spatial information are largely dependent on the size of the study regions. GIS files are very storage consuming. Unnecessary large or detailed information will not only occupy the disk space, but also quickly deteriorate disk access performance as the disk becomes full. On the other hand, decreasing map scale results in lower map resolution as features are smoothed, simplified, aggregated, eliminated, and reduced in dimensions. The choice of scale is possibly the most cost-sensitive database design decision. Secondly, a well-defined study region can eliminate unnecessary information that is irrelevant to the users' needs and all of the data sets can have a consistent geographic extent. This is especially important if the data sets are presented in map layers.

Define the map projection

Map projection is another important element in spatial database design. Earth's spherical coordinates are expressed in terms of latitude (parallels) and longitude (meridians) with a three-dimensional surface, however, the Cartesian coordinate system used in mapping is based on a flat two-dimensional surface. This system is the result of projecting coordinates from the three-dimensional, spherical surface to a two-dimensional flat surface. Real world locations are measured using x-and y-coordinate values from a specified point of origin. The conversion of geographic locations from a spherical coordinate system to a two-dimensional coordinate system causes the distortion of one or more of the spatial properties (area, shape, distance, or direction). A specific map projection can preserve one property at the expense of the others. A map projection is a systematic representation of all or part of the surface of the earth on a plane with controlled or predictable distortion results. To effectively representing the map using spatial data, an appropriate map projection with the least distortion should be chosen.

All of the spatial data sets should be projected into the same projection system from their original projections, before they can be overlaid and analyzed in the same analysis environment.

Phase II Physical Design

The elements in the physical design phase are acquiring data sets based on the users' needs, processing the data sets to a consistent format with appropriate geographic extent and projection, developing the data documentation and metadata.

Acquire the data sets

The cost of data acquisition can be high, especially for the direct creation or automation of the data. Making use of readily available data from the Internet can be very cost effective. Most regional geographic data sets over the Internet are categorized in tiles based on the political boundaries or map quadrangles. To cover the whole study regions, more than one data tile may be needed. All of the data set tiles that cover the study region need to be identified before the data are downloaded from the web sites. Many of the data sites provide the option of selecting the data using graphic, which makes this identification process easier.

Process the data sets

Detailed data processing procedures can be very different depending on the various data types provided by various organizations, however, the processes usually includes the following basic steps if the ESRI spatial data processing software ARC/INFO is employed:

- Unzip the compressed files
- Reformat the data file to be compatible with ARC/INFO
- Merge the data set tiles to form larger continuous seamless data sets.
- Customize the attributes information to meet the user needs
- Project the merged data sets to the desired projection.
- Clip the data sets for the defined study region from the larger data sets.

Additional steps may be necessary depending on the specific data types. For example, the tiles of land use files may not match well at the edges due to individual mapping errors. An edge matching may be necessary after merging all of the tiles together. Sometime there may be a need to add in more detailed attribute information to existing feature attribute tables using local data. In this case, the table editing procedures will be involved.

In a later section of this report, a preliminary design of the Pantex spatial database will be documented, which includes detailed procedures for processing each data type in the spatial database.

Document the data dictionary

The data dictionary describes the structure of the database and any other information that would help users to understand the database. The data dictionary provides the data names, data content, data types (coverage, grid, image, etc.) and data projections (if applicable). The data dictionary may also include the following information (ESRI Educational Services, 1996b) :

- **Lineage** Lineage describes the source materials used and their dates, as well as the processing steps and transformations leading to a final product. This is required for users to determine fitness for use or to enable the maintenance and updating of the data
- **Positional accuracy** Positional accuracy describes the reliability of positional information.

- **Attribute accuracy** Attribute accuracy describes the expected errors of omission and commission in identifying features to be mapped, along with their characteristics (attributes).

Prepare the metadata document

Metadata are frequently referred to as data about data. They are additional information that is needed for the data to be useful. Metadata are comprehensive, systematic, and deductive information about the content, the structure, the relationships, the representation, and the use context of the data stored in the underlying database. Metadata management is multidimensional, it includes the following aspects (ESRI White Paper Series, 1997a):

- Inventorying existing data holdings
- Defining the names and data items to facilitate understanding
- Building a keyword list of names and definitions
- Indexing the inventory and the keyword list for access
- Recording processing steps performed on the data including those involved in the initial collection
- Documenting the data structure used and the data model implemented
- Recording the logical and the physical database schema
- Documenting the relationships between data items in different data sets as well as between individual data sets

- Recording the processing steps performed on the data including precollection decisions, collection methods, conversion, and postconversion editing and analyses
- Documenting the representation (such as map projections) chosen for the data
- Documenting application-specific metadata including flowcharts for macro language programs
- Updating the metadatabase in a consistent fashion and at regular intervals.

3.2.2 Spatial Database Management

The spatial database management includes database administration and database distribution. (Figure 3.1)

3.2.2.1 Spatial Database Administration

Data base administration involves the daily database storage and daily database maintenance, such as update and backup.

Database storage

Currently, the spatial data can be stored in five different formats using ESRI software. These five storage formats are listed below (ESRI Educational Services, 1996c):

ARC/INFO LIBRARIAN If the spatial database logically manages as a set of map sheets, ARC/INFO LIBRARIAN can be used to physically implement this system as a map library. ARC/INFO LIBRARIAN manages large coverages by breaking them up into tiles. Each tile is physically a workspace that stores all the coverage data for a give spatial extent. ARC/INFO LIBRARIAN is a transactional map library, it has a mechanism for keeping track of which user is working on which map section, so that two users do not interfere with each other during editing operations. It can also partially track the historical information (i.e., it keeps a history of who modified which map section, but there is no specific historical information about exactly what changes were made.

ArcStorm If the data management is feature-based or if complex transactions involving multiple table and/or features must be managed in a consistent way, an ArcStorm database is the appropriate choice. ArcStorm allows multiple users to be working in the same coverage or table at the same time, but not the same feature or row. ArcStorm also supports full historical tracking. The transaction history can be optionally stored and used to roll back the database or to create historical views of the data.

Spatial Database Engine (SDE) If database contain millions of features and they require a standard relational database environment in a client/server, cooperative processing architecture, the SDE system is the appropriate choice. SDE is a set of data access and geoprocessing tools which provide programming level access to the geographic applications. The data model is feature-oriented, which means that geometric and attribute data pertaining to a spatial feature (e.g., a land parcel) are stored together and require only one disk access to retrieve. Spatial features are stored in one continuous database. In many respects, SDE is an ideal geoprocessing server for large Internet database because it offers the following advantages:

- High-performance geoprocessing capabilities
- True Client/server design
- All the advantages of commercial relational database management systems (backup, recovery, networking capabilities, tools etc.)
- Multiplatform, multiprocessor support
- Open applications interface for programmer interaction
- GIS clients with in-built Internet capabilities
- Supports large numbers of both read and write concurrent transactional users.

SDE does not have a graphical user interface. The ability to write computer programs and use a C application programming interface (API) are required to use SDE.

Workspaces If a database contains limited spatial layers with manageable extent, and does not require SQL queries, multi-user transaction, and historical tracking, a workspace database management system is sufficient. A workspace is a directory that can contain coverages, grids, and INFO files. It can also contain any other kind of files or directories. The INFO files lack the nice features listed above, but provide a powerful and easy analysis environment. ARC/INFO provides utilities for converting tables between INFO, SQL, and dBASE formats.

External Relational Database Management System An external relational database management system can also be chosen to store tables. The tables can be accessed using the DATABASE INTEGRATOR, ArcStorm databases, or via ArcView. One of the reasons to choose the external relational database management system is that a significant amount of data and tables have been stored in an external relational database and the spatial information is not the center of the database. In this case, the queries can be performed in the external relational database to retrieve the subset of data and GIS software can be used as analysis tools.

Database maintenance

Database daily maintenance mainly involves database update and backup. To design a well organized, comprehensive and high quality spatial database requires a tremendous effort. To keep it accurate and updated requires no less effort. A well-coordinated database update schedule should be developed for different types of data. Many environmental operations and sampling data need daily update; some others may need weekly, monthly or annually update. Regional geographic data may not need to be updated as frequently as environmental operation and sampling data, however, the database administrator should make a effort to keep the data in reasonably recent state by coordinating vendor support, input/output control and performance testing.

Data backup is one of the most crucial system administration tasks. A regular backup schedule should also be developed and implemented. Back up ensures file system integrity in case of a system failure and protects GIS databases in case they are deleted accidentally. System security can also be broken by natural disasters such as fire, flood, earthquakes and hurricanes. Several companies provide off-site backup services where organizations pay a fee for the safeguarding of GIS database and project backups. These companies have their own security systems and are built to withstand natural disasters. For a list of companies that provide system backups services and a detailed list of backup software packages see RS/Magazine, January 1996 and RS/The Power PC Magazine, December 1995. (ESRI Educational Services, 1996d)

3.2.2.1 Spatial Database Distribution

The effectiveness of a spatial database is largely dependent on the accessibility of the database, defined by the following criteria:

- Database distribution range – How many people can access the database?
- Database distribution format – Who can access the database with what kind of the requirement?
- Easy of use – How fast is the data retrieving speed? How easy to query or search the data or features of Interest? What is the file space requirement?
- Data Currency – How often is the database being updated?

There are several options for distributing the spatial database to end-users. The most frequently used options are listed below:

- Online database

- CD-ROM
- Internet/Intranet

Online database

An online spatial database has a limited distribution range. The primary disadvantage of an online spatial database is that only users having access authority to the network can access the data. Another disadvantage is the operational dependence on the access through networks such as wide area networks (WAN) and local area networks (LAN) to the centralized database. Access to database is lost when network communications fail between the user and the central database.

The distribution of an online database may also be limited by the network bandwidth. The spatial data are very graphic intensive, with large file sizes. Heavy load of the inquiries may slow down the processing and make it difficult to use the spatial data. This situation should not be a serious problem for the users who connect to the spatial database using Local Area Network. Local Area Network provides high bandwidth communication within relatively short distances, while Wide Area Network provides low bandwidth communications over long distances. Therefore the speed of retrieval of the spatial data from an online spatial database using a Local Area Network should be faster than the retrieval through a Wide Area Network and Internet/Intranet but slower than direct retrieval from local disk space.

The primary advantage of an online database is that all data can be managed in a single environment. Data can be easily updated and backed up. An online database is especially useful for managing a large database, which has data in the gigabyte or terabyte range. Only one copy of data is maintained. Users only retrieve a subset of the data for an individual application. Therefore the storage space is saved and data integrity is preserved. Another advantage is that if the spatial data is directly stored in a well-defined relational database environment, topologically constrained searches and query are allowed, which provides exceptionally rapid spatial data retrieval and dynamic spatial overlay on the database.

The implementation of an online spatial database requires complicated system design, construction and maintenance. The cost of hardware and software could be significant depending on the size of the spatial database.

An online database is usually maintained and updated daily by a database administrator. It provides the most accurate and updated spatial data.

CD-ROM

CD-ROM is another very popular database distribution option.

The CD-ROM publishing industry started booming in 1990s. CD-ROM’s have become a standard media for software and data publishing. Low priced CD-ROM mastering technology has made CD-ROM publishing affordable for almost any institution. The millions of CD-ROM drive attached to desktop computers, and CD-ROM standards such as ISO9660, make CD-ROM data delivery a secure choice. And the 600 MB-plus capacity and low per-unit production cost of a CD-ROM disk make it an ideal medium for delivering large data sets to a broad audience.

The primary advantage of CD-ROMs as an option of spatial database distribution is high speed of data access. The graphic intensive large data sets are retrieved directly from the local disk space without any interference of the network performance. The primary disadvantage of the data file on the CD-ROM is that they may not be the most updated data. Further more, ordering the CD-ROM can be expensive and time consuming.

A CD-ROM is more suitable for regional geographic data sets because of their less need for the data update. CD-ROM as a distribution option for a spatial database is best for relatively static spatial data, because of the more data distributed in the CD-ROM format, the less the unit price of the CDs, and the less burden for the network. The network channels can be saved for more urgent and updated data transactions. The combination of the CD-ROM for basic static information and the Internet for more updated and timely information is likely to be a future trend.

Internet/Intranet

The term *Internet* is usually used for the global encompassing public network. The term *Intranet* is applied to private networks that use standard Internet technologies. An Intranet is usually separated from the Internet by a firewall (an electronic device for filtering network traffic).

The Internet is the world’s largest network, which is unified by common use of the Internet Protocol (IP). This communication standard allows heterogeneous hardware to communicate effectively in a common environment that is open, inexpensive, easy to use, supports multimedia (graphic, video, and sounds) and provides hyperlinking capability.

In the context of GIS, the Internet has many potential uses such as data publishing, product sales and distribution, and GIS services (e.g., on-line geocoding).

Internet/Intranet as an option of spatial database distribution has the primary advantage as having the largest distribution range. More and more government agencies and other data providers are publishing their public domain data over the Internet/Intranet. To obtain these data, it used to take months to process the orders with a significant cost. With the “Information Super Highway - the Internet”, it just takes the moment of clicking the mouse, and most likely the data is free.

Internet/Intranet solutions not only provide the data fast and free, but also provide the data in a more updated manner. With a well-interfaced relational database management system, the Internet/Intranet can even provides spatial data that are updated daily or even dynamically changing data.

Further more, Internet/Intranet solutions not only deliver spatial data online, but also have the potential to deliver the GIS functionality online. The ideal Internet solution for GIS should meet the following requirements (ESRI White Paper Series, 1997b):

Simple standard tool and platform independent The GIS data and processing should operate using standard browser and server technology and be independent of heterogeneous hardware environments.

Dynamic and secure data access Users should be able to access data interactively and publish maps directly. By separating users from the database, integrity and security can be maintained easily without significant overhead

Rich mapping and GIS capability A centrally managed servers that have sophisticated geoprocessing tools for display, query, and analysis can hold and process data and then publish to browser clients, providing both rich mapping and geoprocessing capability.

Wide range of data formats Both raster and vector data should be supported in the Internet applications. Both of the data types may be used for geoprocessing, but raster data can be used for displaying a raster map which is more like a "real map".

The major disadvantage of the Internet/Intranet solution is the heavy dependence on the reliability of the network. Failure of the network can completely cut the access to the database. The heavy traffic over the Internet can also slow down the data access process and make it difficult to use, especially for the large and graphically intensive GIS files.

3.2.3 Spatial Database Implementation and Training

3.2.3.1 Implementation

Spatial database development and management are complicated processes whereby raw data sources, data process, database management methods, quality control procedures, desired analytical operations and data distribution operations are all integrated into a formal plan that provided an integrated spatial analysis environment to wide range of users.

These complicated processes are not successfully implemented without an intimately cooperative effort among the users themselves. A good spatial database developer should have a close relationship with users, understanding them so that the database developed will meet the users' needs closely, teaching them enough about the design process so they can control it directly and independently, providing an overview to help link the different user requirements, and listening to the feedback for improvements.

Several issues need to be considered in the implementation process of a spatial database.

Budgets Budgets supporting the spatial database and the distributed computer system environment are largely dependent on the users' needs and performance expectation, which should be evaluated early during the user assessment stage.

Hardware/Peripherals The hardware/peripherals for a spatial database and it related application are computers (Mainframe, workstations, PC, Macintosh), graphics terminals (X-terminals, X-terminal emulators), digitizers, scanners, plotters, printers and storage devices. The most important issue of hardware is the compatibility. Never buy any hardware or peripheral equipment without making absolutely sure that it will work correctly with your existing hardware and software.

Operating systems Many operating systems such as UNIX, Open VMS, Windows, Windows95, and Windows NT, Macintosh and Power Macintosh, DOS, can be used in spatial database development and management. Unix and Windows NT are the two most popular ones. ARC/INFO, ESRI's major GIS geoprocessing and data management software, used to work only on UNIX, making UNIX the essential operating system for GIS. However, with the presence of PC ARC/INFO and the Window based software such as MapObjects, the Windows NT operating system becomes more and more popular. Also ARC/INFO itself now works on Window NT.

Networks Networks can be used for electronic mail, file transfers, remote logins to other workstations, and remote disk mounting for online access to any disk in the network. Remote disk mounting is a technique for sharing data across the network. A spatial database can be stored on a server, while users can retrieve the data by mounting the server disks on their own computer. A common alternative to network file system (NFS) communications is Client/Server communications. This requires a communication API (Application Program Interface) installed on the client machine and a communication server process on the server to support the communications. Client/Server communication is much more efficient than NFS.

3.2.3.2 Training

A critical component of successful spatial database implementation is people. GIS training and software training are the key aspects of a successful operation and application of the spatial database. The training should be an ongoing program because of the rapid progress of GIS technology. The following is a list of a few options (ESRI Educational Services, 1996e) that can be employed for an ongoing GIS training program.

On-site training Regularly organizing the people that want to have training on GIS or other database software is the most direct route for an ongoing training program, and it is also more cost effective, provided the computer facility is available.

Customized training Customized training associates the spatial database with the work of the users and tailors the course content expressly for the uses, with a focus on the procedures they will use most.

Online tutorial Nowadays online tutorials are far more advanced that they were used to be, providing the users with a very explicit, easy-to-follow environment for self-study.

Internet/Intranet communication One of the most effective ways of training is the cross training by the users themselves. Research and application Internet/Intranet provides the ideal environment to share the information related to the newest developments, the potential problems and difficulties, the tips and suggestions for the better use of the spatial database, and the application products using the newest development. Learning GIS over the Internet will be one of the future trends. ESRI has lunched the education program Virtual Campus to provide the GIS training over the Internet. GIS exercises on the homepage of Dr. David Maidment provide excellent training materials on GIS in hydrology.

The above sessions have introduced the concepts of the Environmental Information System and the framework of Environmental Information System. Each component in the framework has also been discussed in detail. The following sessions will introduce some preliminary investigations on some of the components of the Environmental Information System framework as application to the Pantex plant.

Chapter 4: SPATIAL DATABASE DEVELOPMENT FOR PANTEX ENVIRONMENTAL INFORMATION SYSTEM

The spatial database development of Pantex and the surrounding area is part of the project, Risk Characterization of Plutonium Mixed-Oxide (MOX) Fuel Facility at the Pantex Plant. The objective of that project is to develop a spatial database framework for assisting in characterization of environmental risks from a MOX fuel processing facility at the Pantex Plant, and to develop and analyze potential pathways for human and ecological exposures. Through development of the database, information gaps will be identified so that additional research may be initiated in a timely fashion, and informed decisions may ultimately be made.

The direct users for this database are the Risk Characterization project team. The major issues in the Risk Characterization project are shown in Figure 4.1.

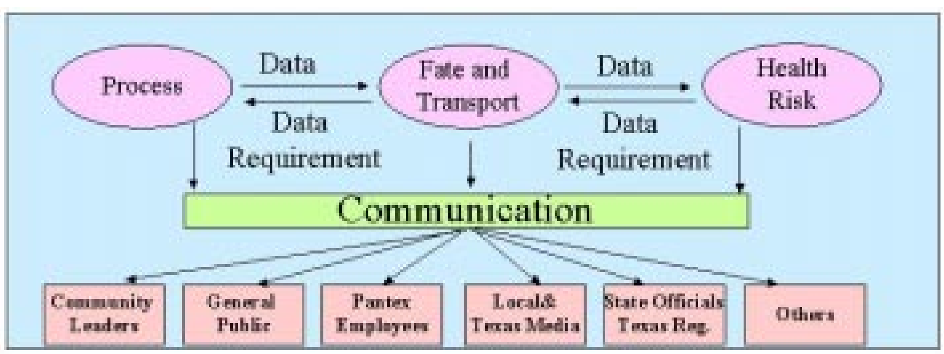


Figure 4.1 Technical Subjects of Risk Characterization of Pantex MOX Fuel Facility

For the fate and transport issues, environmental characterization of the Pantex facility and its surrounding region is necessary. This characterization includes surface description such as streams, land use, vegetation and soils distributions and subsurface description such as the groundwater distribution. For health risk issues, census data are necessary. Communication is extremely important for this project; a very explicit description of the Pantex facility and its surrounding region is necessary. In addition, an open and easy-to-use channel should be set up to provide the environmental information to the public.

After a detailed user assessment, the following eight types of environmental data were included in the regional spatial database:

- Raster Image of the Pantex site and surrounding region
- Census description of population
- Stream and Transportation Digital Line Graph
- Land use distribution
- Soils distribution
- Vegetation distribution
- Aquifer distribution
- Facility layout

Most of the above information was researched and retrieved from the Internet. The raster image data were purchased from Horizons Technology, Inc. The facility layout in CAD drawings were provided by Pantex GIS specialist Gary Thomas.

The initial study region was required to be 50 miles radius range, centered on the Pantex facility. A 100 x 100 miles square centered on the Pantex facility were decided later for the convenience of mapping. Figure 4.2 shows the study region and the coordinates used for this spatial database. In the raster image of Figure 4.2, Pantex is located in the center of the image map, 17 miles north east of the city of Amarillo (the yellow area in the middle of the map). Lake Meredith (the blue area) is located to the north of the Plant. The total study region covers 20 counties. The names and the Fips code of the twenty counties are listed in Table 4.1

The State plane coordinates list below were chosen to minimize the distortion of the projection:

```
Projection STATEPLANE
Fipszone 4201
Datum NAD83
Zunits NO
Units FEET
Spheroid GRS1980
Xshift 0.0000000000
Yshift 0.0000000000
Parameters
```

This projection coordinates is also consistent with that of the Pantex facility CAD drawing files. Therefore, the CAD drawings can directly overlay on top of the regional layouts.

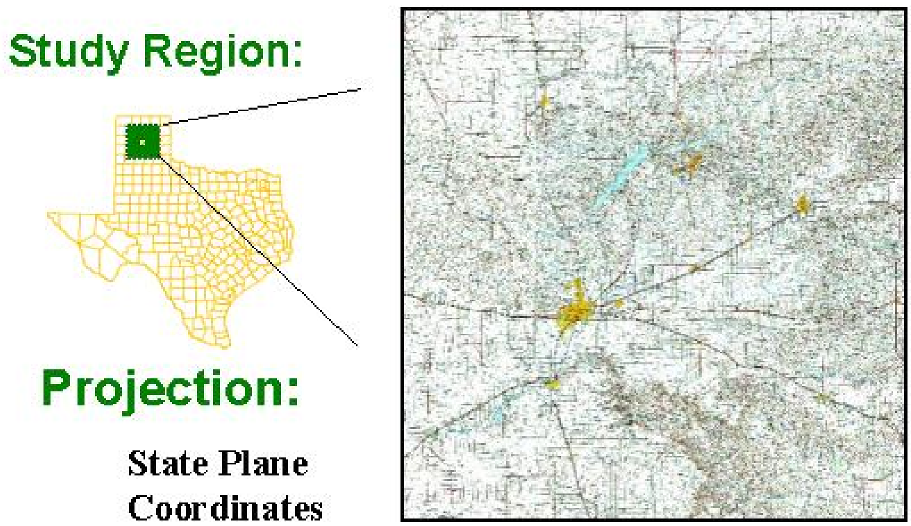


Figure 4.2 The Study Region and the Projection of the Spatial Database

Table 4.1 Names of the Twenty Counties in the Study Region and The Fips Codes

County Name	Fips Code	County Name	Fips Code
Armstrong	011	Briscoe	045
Carson	065	Castro	069
Dallam	111	Deaf Smith	117
Donley	129	Gray	179
Hall	191	Hansford	195
Hartley	205	Hutchinson	233
Moore	341	Ochil tree	357
Oldham	359	Potter	375
Randall	381	Roberts	393
Sherman	421	Swisher	437

In the following sections, the data acquisition and process information are documented for each type of data.

4.1 RASTER IMAGE DATA

The raster image is a scanned map at a specific scale. The purpose of including this type of data into the spatial database is to provide the users the geographic context of the spatial database, so that the users can easily relate the spatial data with the map information.

The data requirement of the image map is a full color, accurately georeferenced digital raster map. In this spatial database, the Sure!Maps Raster of Horizons Technology, Inc. (http://www.horizons.com) located in San Diego was chosen because of its moderate resolution and cost. The 1:250,000-scale digital map was chosen for the study region. It provides reasonable resolution with relatively small file storage requirement. A 1:100,000-scale or 1:24,000-scale map provide better resolutions, however, the file size is too large to be handled for the defined study region. The following section provides the details for retrieving and processing the raster image data set for the Pantex spatial database.

4.1.1 Retrieving the Image from Sure!Map Raster

A piece of the raster image with a range of the desired study region was retrieved from the Sure!Map Raster 1:250,000-Scale CD with the utility software provided by Horizons Technology, Inc.

4.1.2 Converting the Image to a Grid and Projecting to State Plane Coordinates

The initial image was in geographic coordinates. It needed to be projected into State Plane Coordinates in order to be consistent with the other data sets. However, the image cannot be projected directly, it has to be converted to a grid before it can be projected. An Arc command *Imagegrid* was used to convert the image to a grid. A *colormap_file* is used to store the color information when an image is converted to a grid.

```
Arc: imagegrid map250.tif grid250 color250
Converting Image to Grid ...
```

in which the *map 250.tif* is the initial image, the *grid250* is the converted grid, and the *color250* is the *colormap_file,* which describes the color information of the raster map.

```
Arc: project grid grid250 grid250sp geoddsp
Project...
```

in which *grid250sp* is the output grid in State Plane Coordinates, *geoddsp* is the projection file that converts the geographic coordinates to State Plane Coordinates.

<Note: In the above command, the projection file geoddsp is assumed to be in the current directory. Otherwise the path to the projection file needs to be specified.>

4.1.3 Converting the projected grid back to an image

After the projection, the grid in State Plane Coordinate needed to be converted back to an image. An Arc command *gridimage* is used to convert the grid to image. The colormap_file *color250* produced previously was used to describe the color of the image.

```
Arc: gridimage grid250sp color250 map250sp TIFF
Converting Grid to Image ...
```

in which *map250sp* is the resulting image map in State Plane Coordinate, Tiff indicates that the image type is TIFF image. The study region image displayed in Figure 4.2 is the raster image map.

4.2 CENSUS DATA

Census data were acquired from Texas Natural Resources Information System (TNRIS). Two different data sources are available from the TNRIS ftp site. TIGER data produced by the Census Bureau and USGS consists of 1990 decennial census data at the tract level with the geographic boundaries for each of those tracts. The TIGER census data for Texas provides census information for the Texas in block level. An average of five blocks make up one tract, therefore data at the block level is more detailed geographically. However the TIGER census data for Texas is not attributed at the block level, the data are initially contained in a separate block level database. In order to achieve the highest resolution of census data, the two sources of data had to be joined together to provide the census information at the block level with geographic entities. A unique field, identical in TIGER and the block level database, had to be created to make this join possible.

4.2.1 Data Acquisition

Both the TIGER data and the Texas Census data in block level were downloaded from the TIGER92 directory (ftp://www.tnris.state.tx.us/pub/TIGER92/) of TNRIS ftp site. The TIGER data are located in the *tx* directory, while the Texas Census data are located in the *Support* directory as a dBase file *txcensus.dbf*.

The TIGER files are listed by fips code for each county. The Fips code index (ftp://www.tnris.state.tx.us/pub/gis/census/documents/fipscode.asc) is also listed on the web site.

In order to cover the study region, twenty counties' TIGER census data needs to be downloaded. The name of the twenty counties and their Fips codes are listed in Table 1.

The TIGER data at these locations are not presented in an interchangeable .e00 format, but rather in an open coverage format. The files needed for this spatial database are located in both *cengeo* and *info* directory. For each of the counties listed above, both *cengeo* and *info* directories were downloaded and stored in a directory with the county name. <*Note: if UNIX ARC/INFO will be used to process the data, make sure the directory and file name are in lower case.*> These data can be viewed in ArcView as a coverage of census tracts for each counties.

These coverages can then be merged using ARC/INFO command *mapjoin*. The syntax for *mapjoin* is the following:

MAPJOIN <out_cover> {feature_class...feature_class | template_cover} {NONE | FEATURES | TICS | ALL} {clip_cover}

In merging the census data, one of the coverages (./armstrong/cengeo) is used as the template_cover, and all of the features described in this coverage will be joined.

```
Arc: mapjoin bigcensus ./armstrong/cengeo features
Enter Coverages to be MAPJOINed (Type END or a blank line when
done):
=================================================================
====

Enter the 1st coverage: ./armstrong/cengeo
Enter the 2nd coverage: ./castro/cengeo
Enter the 3rd coverage: ./conley/cengeo
……
```

The coverages of the twenty counties were joined to form a large census coverage *bigcensus*. The coverage needs to be cleaned and built for the correct topology.

```
Arc: clean bigcensus
Arc: build bigcensus
```

The Texas Census data *txcensus.dbf* contains census data for all Texas counties in the block level with a total of 15664 blocks. Only the data from the twenty counties listed above needs to be retrieved from the *txcensus.dbf*. A query builder in Arcview3 was used to select the data for twenty counties. <*Note: the cnty column contains characters instead of numbers, so quotation marks need to be used in the query builder.* > The query in the scratch pad should be similar to the following:

```
([Cnty]  = "011”) or ([Cnty] = "045") or ([Cnty] = "065") or ……
```

The selected data for all twenty counties were then exported to dBase file *bigcenbk.dbf* (Census data for the twenty counties at the block level). An Arc Command *dbaseinfo* was used to convert dBase file *bigcenbk.dbf* to Info file *bigcenbk.*

```
Arc: dbaseinfo bigcenbk.dbf bigcenbk
dBASE table bigcenbk.dbf copied to INFO table bigcenbk
Items: 40, Records: 407
```

Bigcenbk is the Texas Census table in the Info table format that contains counties census data in the 407 blocks, covering 20 counties.

4.2.2 Data Processing

To display census data at the block level, a unique field, identical in TIGER data table and Texas Census data table, had to be created so that the two could be joined. The field BLCKGR-ID in TIGER data table was chosen as the unique field. BLCKGR-ID is a concatenation of the fields CNTY, CTBNA, and BLK in the TIGER data table. You can view these three fields of the attribute table of theme *region.pop*. This field was chosen because the same field could be created in the Texas Census table by combining the fields CNTY, TRACTBNA, and BLCKGR in Texas Census table. To be identical in the two tables, BLCKGR-ID should consist a total of 10 characters contributed from the three fields as indicates Table 2. However, the way in which TIGER joined the fields to create BLCKGR-ID is different from that of Texas Census data.

Table 4.2 Composition of BLCKGR-ID in TIGER Data and Texas Census Table.

	BLCKGR-ID(10)		
TIGER Data Table	Cnty(3)	CTBNA(6)	BLK(1)
Texas Census Table	Cnty(3)	TRACTBNA(6)	BLCKGR(1)

In TIGER data table, when CNTY consists of only 2 characters the BlCKGR-ID will consist of only 9 characters, omitting the first zero digit. Since BLCKGR-ID should consist of 10 characters, the first digit zero needs to be added if necessary. In Census, however, the missing digits are usually the two zeros at the end of field TRACTBNA, when it only has four characters. Therefore two zeroes have to be appended to the end of TRACTBNA if the field only has 4 characters, before it is used in the creation of BLCKGR-ID.

The table will be edited in ARC/INFO Table module.

```
Arc: tables
Enter Command: dir
```

A list of external tables with their related internal names will show up. The Texas Census data for the twenty counties *Bigcenbk* and the TIGER data for twenty counties *bigcensus.pat* (polygon attribute table) should be two of them.

4.2.2.1 Edit Texas Census data table – append two zeros at the end of field TRACTBNA

```
Enter Command: select bigcenbk <Select the table>
407 Records Selected.
```

The Table command *items* can be used to view the structure of the table. The following is the description of the first few columns that are related to editing the field *TRACTBNA*.

```
Enter Command: items

COLUMN   ITEM NAME        WIDTH OUTPUT  TYPE N.DEC
    1  SUMLEV                 3     3     C      -
    4  STATEFP                2     2     C      -
    6  CNTY                   3     3     C      -
    9  TRACTBNA               6     6     C      -
   15  BLCKGR                 1     1     C      -
```

The field *TRACTBNA* starts at column 9 and ends at column 14. The two potentially missing zeros should be in column 13 and 14. Therefore two zeros should be added into column 13 and 14 if the field *TRACTBNA* initially only has four characters.

To identify the records that have only 4 characters, two temporary items c13 and c14 (stands for column 13 and 14) were created using the *redefine* command in Tables. The two items were then referenced to select all of the records that have blank spaces at column 13 and 14 and to move the zeros to column 13 and 14 of these records.

```
Enter Command: redefine <Creating a new item from existing ones>
Enter Starting Column: 13
Item Name: c13 <temporal item name>
Item Width: 1
Item Output Width: 1
Item Type: c <character display>

Enter Starting Column: 14
Item Name: c14
Item Width: 1
Item Output Width: 1
Item Type: c

Enter Starting Column: <Carriage return to exit this dialogue>
```

```
Enter Command: reselect for c13 = '' <select all of the record
                                      that is    blank in c13>
369 Records Selected.

Enter Command: move '0' to c13

Enter Command: aselect <unselect c13, all records now present>
407 Records Selected.

Enter Command: reselect for c14 = ''
369 Records Selected.

Enter Command: move '0' to c14

Enter Command: aselect
407 Records Selected.
```

Now all records in TRACTBNA should contain 6 characters and the unique field can be created. A new item *id* was created from existing fields using the Arc command *Redefine*. The new item needs to begin at column 6 and go through column 15 (10 columns wide) to cover CNTY, TRACTBNA, and BLCKGR. Another new field BLCKGR-ID was added to the end of the Texas Census table *bigcenbk*. The values of BLCKGR-ID were defined by item *id* using Tables command *move*.

```
Enter Command: redefine
Enter Starting Column: 6
Item Name: id
Item Width: 10
Item Output Width: 10
Item Type: c

Enter Starting Column:

Enter Command: sel <Select the table bigcenbk>
File BIGCENBK is now closed.

Enter Command: additem bigcenbk blckgr-id 10 10 c

Enter Command: sel bigcenbk
407 Records Selected.
```

```
Enter Command: move id to blckgr-id
```

4.2.2.2 Edit TIGER data table – add zeros at the beginning of field Blckgr-id if the records only have nine characters.

```
Enter Command: select bigcensus.pat
19361 Records Selected.
```

The following is the description of the few columns that are related to editing the field *BLCKGR-ID*.

```
Enter Command: items

COLUMN   ITEM NAME        WIDTH OUTPUT  TYPE N.DEC
……
……
   19  COUNTY               3     3     I      -
   22  CTBNA                6     6     I      -
   28  BLK                  4     4     C      -
……
   86  BLCKGR-ID           10    10     I      -
```

The temporary item *c86* based on the first character of the field BLCKGR-ID was created using Tables command *redefine*. If *c86* is blank, a zero will be added into the column86 using Tables command *move*.

```
Enter Command: redefine <creating a new item from the existing
                        field>
Enter Starting Column: 86 <The field Blckgr-id start at column86>
Item Name: c86 <a temporal item name>
Item Width: 1
Item Output Width: 1
Item Type: c

Enter Starting Column: <Carriage return to exit this dialogue>

Enter Command: reselect for c86 = ''
3611 Records Selected.

Enter Command: move '0' to c86
```

```
Enter Command: aselect
19361 Records Selected.
```

To display the TIGER data at the block level, the TIGER data coverage needs to be dissolved by the unique identifier BLCKGR-ID. The field that the coverage is being dissolved by must be a character type. This can be checked using the *items* Command.

```
Enter Command: items

COLUMN   ITEM NAME        WIDTH OUTPUT  TYPE N.DEC
……
86  BLCKGR-ID             10    10     I      -
```

The field type for BLCKGR-ID is I (integer). It can be changed by Tables command *alter*.

```
Enter Command: alter
Enter Item Name: blckgr-id
COLUMN   ITEM NAME        WIDTH OUTPUT  TYPE N.DEC  ALTERNATE
NAME
   86  BLCKGR-ID             10    10     I      -
Item Name: blckgr-id
Item Output Width: 10
Item Type: c
Alternate Item Name:
COLUMN   ITEM NAME        WIDTH OUTPUT  TYPE N.DEC  ALTERNATE
NAME
   86  BLCKGR-ID             10    10     C      -
Enter Item Name: <Carriage return to exit this dialogue>
```

Now the TIGER census coverage can be dissolved using the unique field *blckgr-id*.

```
Arc: dissolve bigcensus bigcendis blckgr-id
 Dissolving bigcensus by blckgr-id to create bigcendis
 Creating bigcendis.PAT format...
 Creating dissolve table...
 Dissolving...
```

```
Number of Polygons (Input,Output) =      19361        373
Number of Arcs     (Input,Output) =      42016       1002
 Creating bigcendis.PAT...
```

Further more, the *bigcensus* data needs to be projected to the State Plane Coordinates in order to be consistent with other data set, and finally the topology of the coverage needs to be built.

```
Arc: project cover bigcendis bigcenstp
/export/database3/pantex/prjfile/lamstapl

<”/export/database3/pantex/prjfile/” is the path to where the
projection file lamstapl, which converting lambert projection to
State Plane Coordinates, was located>

Arc: build bigcenstp
 Building polygons...
```

Since the TIGER data table and Texas Census data table now have the identical unique field, they can be joined in ArcView. In ArcView, the coverage *bigcensp* was added, and the related attribute table was opened. This attribute table contains the TIGER data in the tract level. The Texas Census data table *bigcenbk* in Info format was also added. It has the census data in block level. Since both of these tables have an identical unique field, the two can be joined and the resulting table can be exported as the TIGER data in the block level. Now the census data can be displayed in block level as it is shown in Figure 4.3.

The *bigcensp* covers the data for 20 counties, which is larger than the study region. The study region was defined on the raster image map. However, the polygon coverage can not be directly clipped using an image. The coverage temperate *cuttersp* was created by converting the *bigcensp* coverage to grid, setting a window with the same size as the raster image, then converting back to

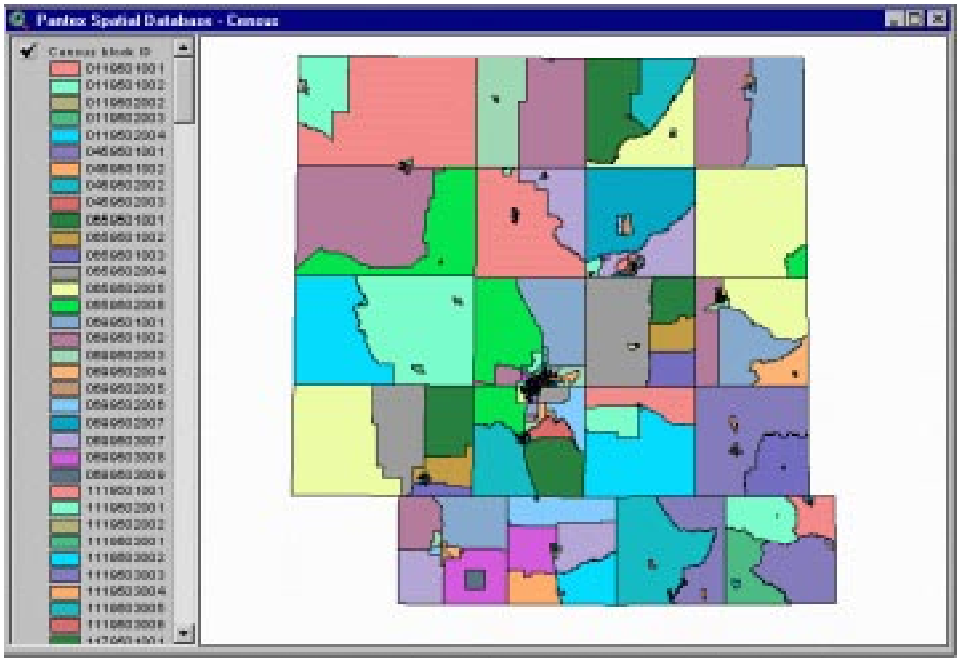


Figure 4.3 Census Data in Block Level for Twenty Counties

coverage. The regional census data coverage is the clipped coverage using template *cuttersp.* The same template is also used define the study region for the other data sets in this spatial database.

```
Arc: polygrid bigcensp bigcenspgr

Grid: display 9999
Grid: mape bigcenspgr
Grid: gridpaint bigcenspgr

Grid: setwindow r_img_gr <r_img_gr is the grid of raster image
                          with the study region. It should be in
                          the current directory, or a path needs
                          to be specified>
Grid: cuttergr = bigcenspgr
```

```
Arc: gridpoly cuttergr cuttersp
Arc: clip bigcensp cuttersp r_cen
Arc: build r_cen
```

The tract layout of the census data for the study region is shown in Figure 4.4.

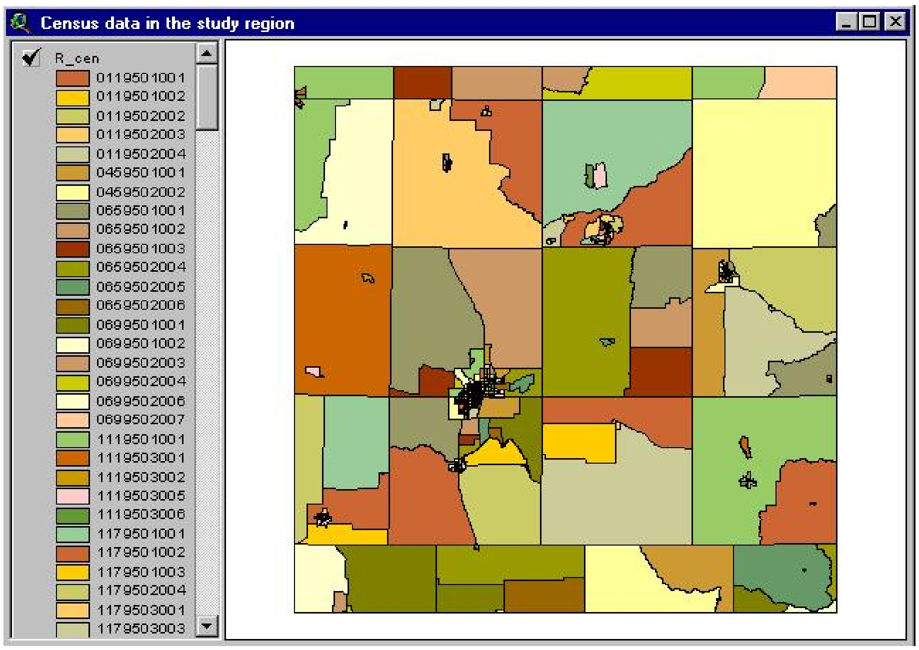


Figure 4.4 Census Data at the Block Level for the Study Region

The Census data provide the information about age and race distribution of the population of the study region. They also provide information such as the household, renter or owner, etc. Figure 4.4.a and Figure 4.4.b give the sample information provided by the Census data. Figure 4.4.a describes the population of children with an age between zero to four. Figure 4.4.b describes the age distribution of the population in the study region.

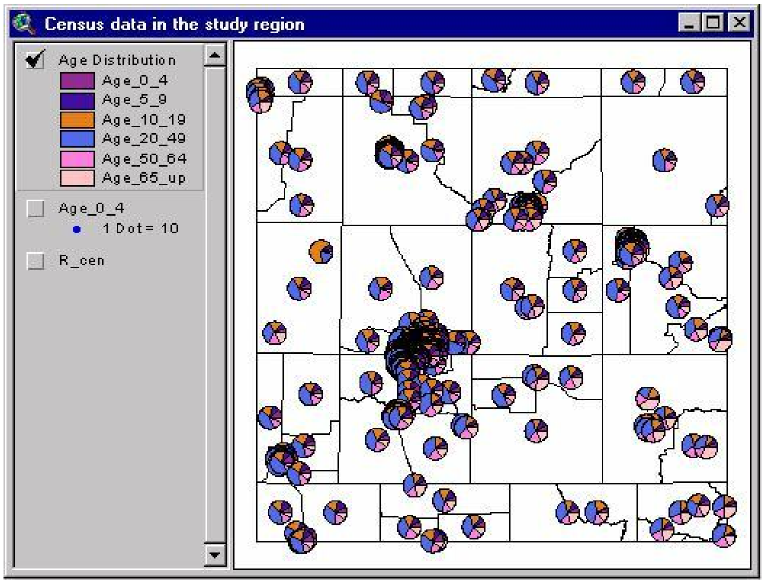


Figure 4.4.a The Population of Children in the Study Region with an Age between Zero to Four

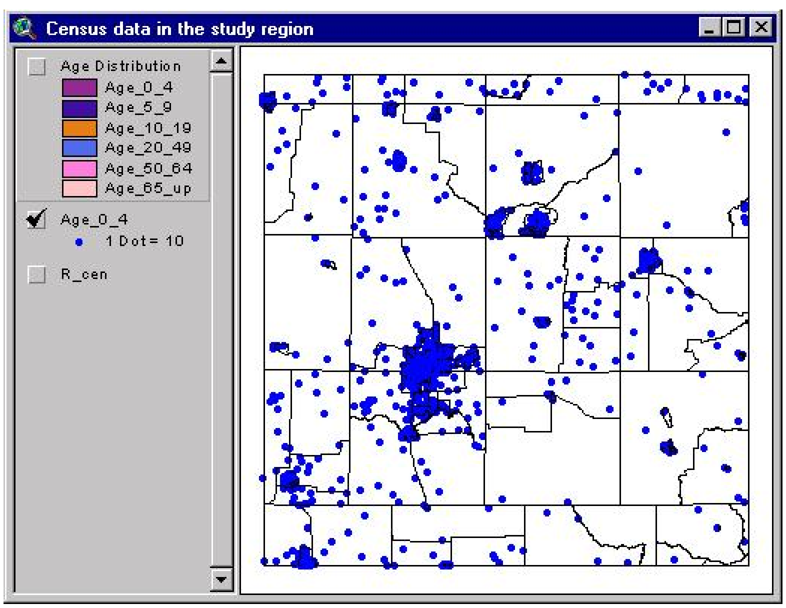


Figure 4.4.b The Age Distribution in the Study Region

It should be noted that although after joining the tables, *bigcensp* theme has the block level information attached, the attribute table of *bigcensp* is not intrinsically modified. In a new view or a new project, the attribute table of *bigcensp* will not have the block level information attached. In order to permanently modify to the attribute table of *bigcensp*, the table was edited by adding the new fields with the same field names as the fields that were temporary joined before, and the values in those fields were copied to the new fields using the field calculator. Finally, the entire temporary joined fields were removed using ArcView function Table/Remove All Join.

4.3 STREAM AND TRANSPORTATION DATA

Stream and Transportation data were processed from USGS 1:100.000 Digital Line Graph data. The source data were retrieved from from CD-ROM "1:100,000-Scale Digital Line Graph (DLG) Data Hydrography and Transportation, US GeoData (Optional Format), August 1993 (Area 8 Texas and Oklahoma)", which was obtained from the USGS Earth Service Information Center in Boston, VA.

The data files for each 1:100,000-scale quadrangle are in individual subdirectories under \100K_DLG. These subdirectories are named using the first eight characters of the quadrangle. There are four files within each of the subdirectories, one for each data layer.

The file names are in the following format: XXXYYYYY.ZIP

In which XXX = Code assigned to the 1:100,000-scale quadrangle. This code and the quadrangle name are in the file FILE.OUT in the DLGSOFT subdirectory.

YYYYY = Layer name: RAIL - Railroads

ROADS - Roads and Trails

MTRAN - Miscellaneous Transportation

HYDRO - Hydrography

ZIP = Extension added by the software to compress the data.

Each of the four files contains the individual data layer files. There are from 4 to 32 files in a data layer depending on the complexity of the data layer.

The data files are stored in a compressed format on the CD-ROM.

The filenames are described in ***USGS-US GeoData site 1:100,000 DLG (http://edcwww.cr.usgs.gov/doc/edchome/ndcdb/ndcdb.html)***, FTP via Graphic

There were a total of twelve 1:100,000-scale quadrangle zipped files used to cover the whole study region for each data layer. Eight of those files were in the UTM projection zone14, while four of them were in UTM Zone13. All of the file names and associated code are listed in Table 4.3.

All of the twelve quadrangles should eventually be merged to form a coverage that covers the study region, however, since they were in two different projection zones, they can not be merge together in the UTM projection. For the development of this database, the eight files in zone 14 were merged and projected to geographic coordinate projection using projection file utmgeo_e (see appendix A); subsequently the four files in zone 13 were merged and projected to geographic coordinate projection using projection file utmgeo_w (see appendix A). The projected files were then merged and projected to a final projection in State Plane Coordinates.

Table 4.3. DLG file Names and Associated Quadrangle Codes

UTM Zone 13		UTM Zone 14	
File Name	Quadrangle code	File Name	Quadrangle code
dalhart	dh4	spearman	py3
hartley	tu2	borger	am1
vega	tu4	amarillo	am3
hereford	xk2	tulia	pv1
		perryton	py4
		pampa	am2
		shamrock	am4
		wellingt	pv2

Each of the zipped files downloaded from the CD-ROM contains eight 15-minute maps. Each 15-minute map file was first converted into an ARC/INFO line coverage using *dlgarc* command. Then, the borders of each of the 15-minute map files are trimmed away from the coverage so that those 15-minute parallels will not appear in the final appended coverage.

After the eight 15-minute map files are trimmed, they were appended together to form a DLG coverage for the corresponding quadrangle. The DLG quadrangle coverages from all eight (for zone14) or four (for zone13) zip files were then merged to form the DLG coverage for the corresponding zone.

These complicated processes were performed on all four layers of the DLG data. An AML (Arc Macro Language) file *dlgmerge.aml* were modified to merge the files in a timely manner. The AML file was modified based on the file types and zones for each application. The modified version is listed in Appendix B(dlg_hydrography is used as example). The original version of *dlgmerge.aml* can be viewed from the appendix B of Bill Saunders' Masters Report "Non-Point Source Pollution Assessment of the San Antonio - Nueces Coastal Basin.

<Note: A file transfered from a PC cannot be recognized by the dlgarc command because the coversion from DOS to UNIX adds a character ^M to each line of the text file. To use the CD-ROM, either use "cp" to copy files at UNIX prompt, or transfer from PC and use dos2unix command: dos2unix filename filename to delete this ^M character. >

The two DLG coverages in UTM Zone13 and Zone 14 were then projected to geographic coordinates, merged, and finally projected to State Plane Coordinates and the topology was built. The regional data were clipped from the larger DLG coverage using *cuttersp* as the template.

```
Arc: clip bighydsp cuttersp r_dyd
```

The streams and transportation data are shown in Figure 4.5 to 4.8. Pantex is located in the center of the region.

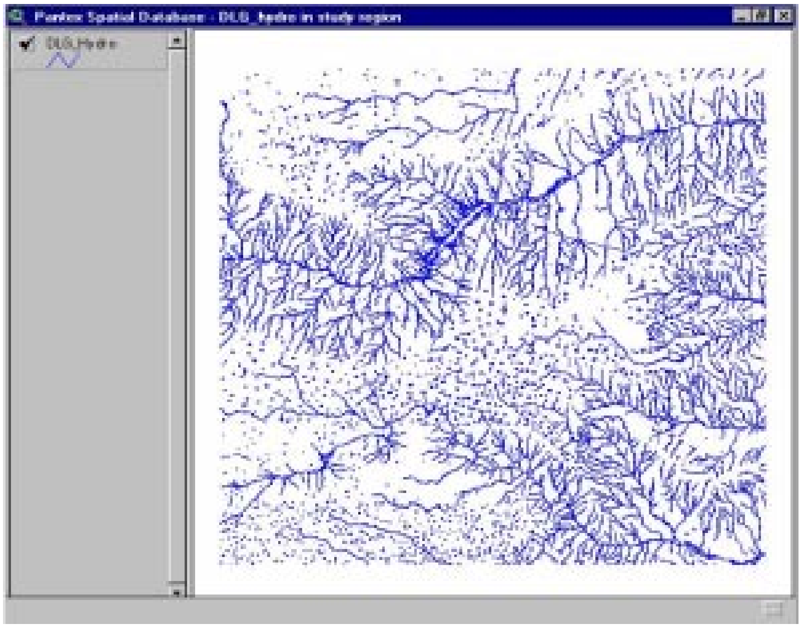


Figure 4.5 Hydrography (DLG) of the Study Region

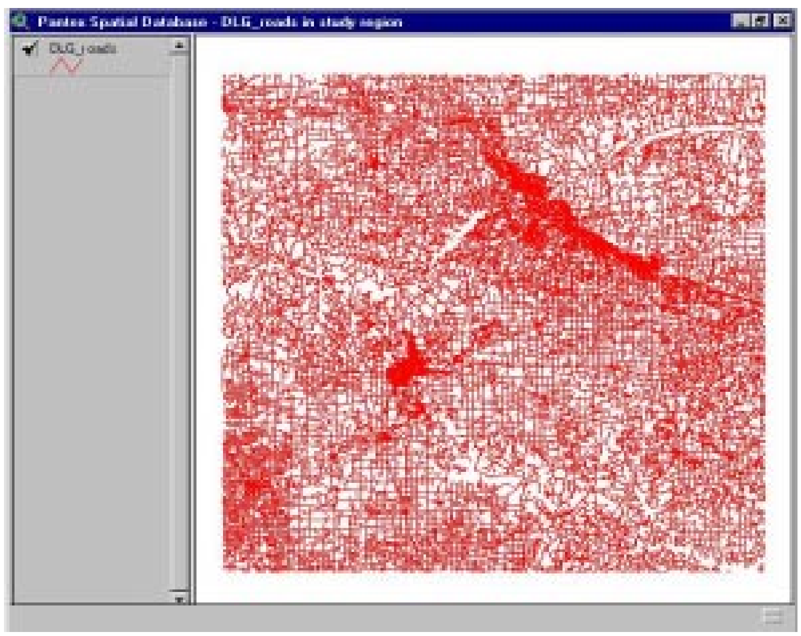


Figure 4.6 Road (DLG) of the Study Region

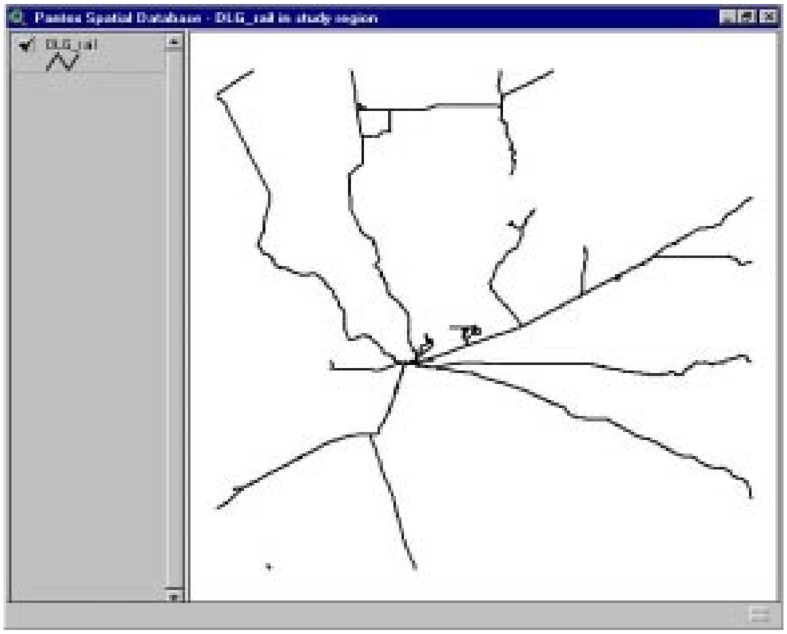


Figure 4.7 Railroad (DLG) of the Study Region

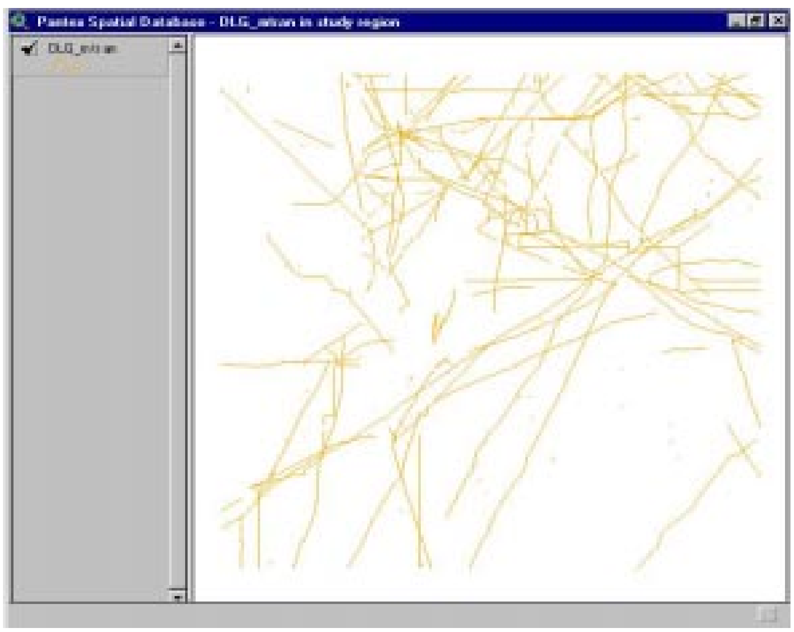


Figure 4.8 Miscellaneous Transportation (e.g. Airport and pipelines)

4.4 LAND USE DATA

4.4.1Retrieving the Source Data

The source data of land use files were obtained from the TNRIS ftp site. The compressed interchange files that can be read in ARC/INFO format are located at ftp://www.tnris.state.tx.us/pub/GIS/topography/LULC/250k_arc/. The names of the required data files can be decided by viewing the ***USGS-US GeoData*** ***(http://edcwww.cr.usgs.gov/doc/edchome/ndcdb/ndcdb.html)*** LULC 250K FTP via Graphic

A total of six land use tiles were used to cover the study region. The six files are listed below:

```
amarilloa3.e00.gz
clovisa3.e00.gz
dalharta3.e00.gz
perrytona3.e00.gz
plnviewa3.e00.gz
tcmcaria.e00.gz
```

The six files were downloaded from TNRIS ftp site to PC, transferred to UNIX using FTP, unzipped using gunzip, imported to ARC/INFO, cleaned and built, and finally dissolved using lulc_code. Using *amarilloa3.e00.gz* as example, the process commands are listed below:

```
$: gunzip amarilloa3.e00.gz
Arc: import cover amarilloa3.e00 amarillo
Arc: clean amarillo
Arc: build amarillo
Arc: dissolve amarillo ama_d lulc_code poly
Arc: lc
Available Coverages
 -------------------
  AMARILLO          AMA_D             CLOVISA           CLO_D
  DALHARTA          DAL_D             PERRYTON          PER_D
  PLNVIEW           PLN_D             TCMCARIA          TCM_D
```

4.4.2 Merging the Data Sets

The six dissolved coverages need to be merged to form a larger land use coverage that covers the study region, however, the coverage CLO_D can not be appended to the other coverages because it has different width for the field of LULC-CODE in its polygon attribute table. This can be checked using Table command *items*.

```
Arc: tables
Enter Command: items clo_d.pat

COLUMN   ITEM NAME        WIDTH OUTPUT  TYPE N.DEC  ALTERNATE
    1  AREA                   4     12     F     3
    5  PERIMETER              4     12     F     3
-
    9  CLO_D#                 4      5     B     -
   13  CLO_D-ID               4      5     B     -
   17  LULC_CODE              3      3     I     -


Enter Command: items ama_d.pat
COLUMN   ITEM NAME        WIDTH OUTPUT  TYPE N.DEC  ALTERNATE
    1  AREA                   4     12     F     3
    5  PERIMETER              4     12     F     3
    9  AMA_D#                 4      5     B     -
   13  AMA_D-ID               4      5     B     -
   17  LULC_CODE              4      4     I     -


Enter Command: items tcm_d.pat

COLUMN   ITEM NAME        WIDTH OUTPUT  TYPE N.DEC  ALTERNATE
    1  AREA                   4     12     F     3
    5  PERIMETER              4     12     F     3
    9  TCM_D#                 4      5     B     -
   13  TCM_D-ID               4      5     B     -
   17  LULC_CODE              4      4     I     -
```

Supposedly the table structure can be changed by Table command *alter*, however, the *alter* command can only change the item name, output and type, it can not change width. To work around this situation, Arcview were used to change the table structure. To do this, start edit the attribute table of clo_d and add a new field with the name: lulc_code, type: number, width: 4, decimal: 0, then use the field calculator to copy the value of original lulc_code to the new defined field and delete the original lulc_code field, finally save the edited result.

The edited coverage clo_d was then merged with other coverages using the command *append*. The resultant coverage was then dissolved by lulc_code.

```
Arc: append biglulam ama_d features <append the following
                                    coverage to form biglulam
                                    using ama_d as feature
                                    template>
Enter Coverages to be APPENDed (Type END or a blank line when
done):
=================================================================
====

Enter the 1st coverage: ama_d
Enter the 2nd coverage: clo_d
Enter the 3rd coverage: dal_d
Enter the 4th coverage: per_d
Enter the 5th coverage: pln_d
Enter the 6th coverage: tcm_d
Enter the 7th coverage:
Done entering coverage names (Y/N)? y
Do you wish to use the above coverages (Y/N)? y

 Appending coverages...

Arc: dissolve biglulam biglulam_d lulc_code poly
```

4.4.3 Correcting Mapping Errors using ArcEdit

Although the six coverages were closely merged, there were still some arcs that were not matched very well. In addition, some of edges were not fully dissolved. These mapping errors can be corrected by *edgematch* command. In this database, these errors were manually corrected using ArcEdit. The major tasks for editing is to remove the extra double or single line between the two merged coverages and to connect the resultant dangling nodes smoothly. The following steps and commands describe a typical process for editing land use coverage using ArcEdit. <*Note: before editing the coverage, make sure to make a copy of original file, in case you make a mistake and delete some important features. Since ArcEdit does not display color, you can view this original copy with color in ArcView. Sometime this helps to identify the features.* >

```
Arc: ae     <start arcedit>
Arcedit: ec biglu_d <edit coverage biglu_d>
Arcedit: display 9999 <activate the arcedit window>
Arcedit: de arcs nodes errors <draw environment: arcs,
                              nodes and errors>
Arcedit: draw

<use ^E to Zoom into the feature that needs to be
corrected>

Arcedit: ef arcs <edit feature: arcs>
Arcedit: sel many

<select the extra arcs by clicking on the arcs that need to
be deleted>

Arcedit: delete

<after the extra arcs being deleted, there would be many
dangling nodes, use move command to connect them>

Arcedit: ef nodes <edit feature: nodes>
Arcedit: move <click on the node 1, type 4 and click on
                              node 2>

<IF two nodes are difficult to be snapped, use editdistance
(ed) to redefine a smaller edit distance. You can also use wt to
change the weed tolerance >

<IF the connection is too sharp, use vertex move to smooth
the angle>
```

```
Arcedit: ef arcs
Arcedit: sel many
```

<select the two arcs wanted to remove the psudo node>

```
Arcedit: unsplit none   <remove the psudo node to make
                        vertex move easier>
Arcedit: vertex move
```

<zoom in and move the vertex to smooth the transition>

<related commands are vertex delete and vertex add>

After finished editing, the *Build* command was used to build the topology of the land use coverage. The coverage was then projected to State Plane Coordinates, built and clipped using the template *cuttersp* to give a regional land use coverage. Figure 4.9 shows the regional land use coverage, using Anderson Land Use Code classification system. Anderson Land Use Code classification system classifies land use types into nine categories listed below:

```
1 = urban
2 = agriculture
3 = rangeland
4 = forest
5 = water
6 = wetlands
7 = barren land
8 = tundra
9 = ice and snow.
```

The second digit distinguishes subcategories of these principal categories,

```
11 = urban residential
12 = urban commercial
13 = urban industrial, etc.
…
```

For more information about Anderson Land Use Code classification system, please refer to corresponding section of USGS Land Use and Land Cover Data Users Guide.

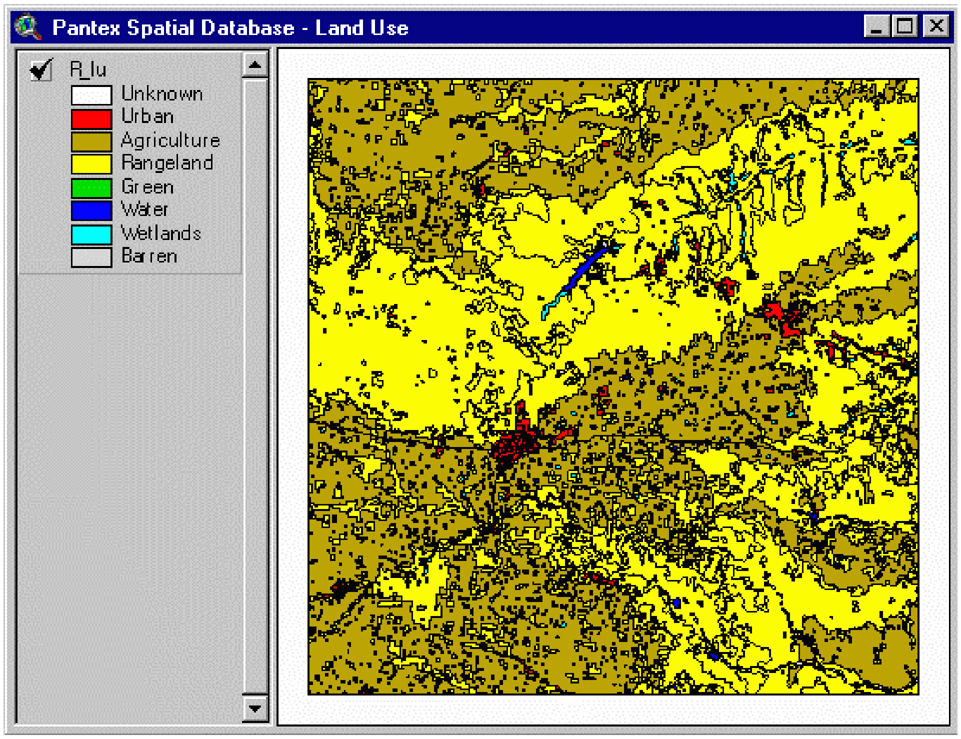


Figure 4.9 Land Use Data in the Study Region

4.5 SOILS DATA

Soils data were processed from STATSGO soil data downloaded from TNRIS ftp site. The file *txsoil.e00.gz* were downloaded from Soil data (STATAGO)(e00 format), (ftp://www.tnris.state.tx.us/pub/GIS/land_use/statsgo/) unzipped and imported to ARC/INFO as a polygon coverage. This coverage contains the 1:250,000-scale soil data for all of Texas. The coverage was projected to State Plane Coordinates the Texas State Map System. The projection file is listed in the Appendix A (*soiltxsp*). The regional STATSGO soil data were clipped from this coverage using the template coverage for the study region *cuttersp*. The STATSGO soil data for the study region is shown is Figure 4.10.

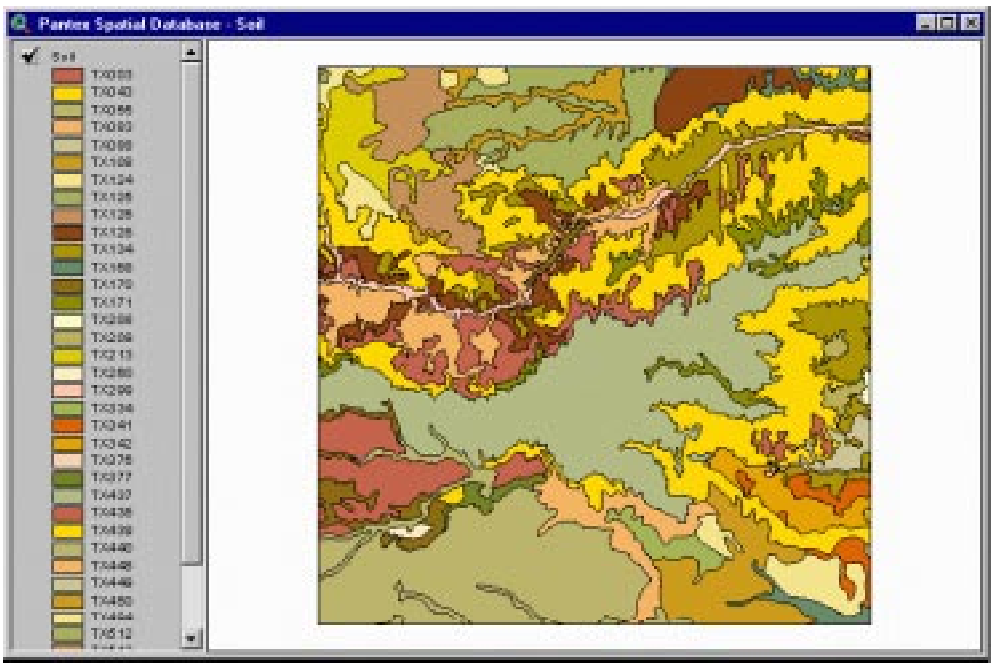


Figure 4.10 STATSGO Soil Data for the Study Region

The descriptive information is provided both in .e00 files and *info* tables. The info tables can be added to ArcView project and linked to the attribute table of the soil data. Therefore the descriptive information of each map units can be viewed.

The descriptive information about the soils data and the soil data of other States are provided in the National Soil Data Access Facility (http://www.statlab.iastate.edu/soils/nsdaf/) provided by USDA Natural Resources Conservation Service Soil Survey Division.

Extensive soil surveys have been conducted in Pantex facility and the surrounding region. These data can also be incorporated into the soil data, however, we have not obtained these data from the facility. This part of database needs to be improved in the future.

4.6 VEGETATION DATA

The vegetation data were processed from two data sources. One is obtained from U.S. Geological Survey - National Mapping Information - EROS Data Center Global Land Cover Characterization web site. http://edcwww.cr.usgs.gov/landdaac/glcc/glcc.html(USGS GLCC). The other one was obtained from TNRIS ftp site vegetation type (e00 format).

4.6.1 USGS LUCC

Global Land Cover Characterization (GLCC) is a 1-km resolution global land cover characteristics database suitable for use in a wide range of environmental research and modeling applications. It is a joint effort of the U.S. Geological Survey (USGS), the University of Nebraska-Lincoln, and European Commission's Joint Research Center (JRC). The global land cover characteristics database is being developed on a continent-by-continent basis. All continental databases share the same map projection (Interrupted Goode Homolosine and Lambert Azimuthal Equal Area), have 1-km nominal spatial resolution, and are based on 1- km Advanced Very High Resolution Radiometer data spanning April 1992 through March 1993. Each database contains unique elements based on the geographic aspects of the specific continent. In addition, a core set of derived thematic maps produced through the aggregation of seasonal land cover regions will be included in each continental database.

Currently the database has data for Global, North America, Eurasia, South America, Africa, and Australia Pacific.

The North America land cover characteristics data are in a flat, headerless raster format. The raster images contain class number values for each pixel that correspond to the appropriate land cover classification scheme legend.

The North America data base is available in two different map projections: the Interrupted Goode Homolosine and the Lambert Azimuthal Equal Area. In this spatial database, the Lambert Azimuthal Equal Area projection system is used because of its compatibility with ARC/INFO.

The grid dimensions of the Lambert Azimuthal Equal Area projection for the North America land cover characteristics data set are 8,996 lines (rows) and 9,223 samples (columns) resulting in a grid size of approximately 83 megabytes for 8-bit (byte) images. The following is a summary of the map projection parameters used for this projection:

```
Projection Type: Lambert Azimuthal Equal Area
Units of Measure: meters
Pixel Size: 1000 meters
Radius of sphere: 6370997 m
Longitude of origin: 100 00 00 W
Latitude of origin: 50 00 00 N
False easting: 0.0
False northing: 0.0
```

```
XY corner coordinates (center of pixel) in projection units
(meters):
Lower left: ( -4486549.963, -4515131.712)
Upper left: ( -4486549.963, 4479868.288)
Upper right: ( 4735450.037, 4479868.288)
Lower right: ( 4735450.037, -4515131.712)
```

In addition to the detailed seasonal land cover regions, the following derived data sets are included in the land cover database:

- Olson Global Ecosystems (Olson and Watts, 1982)
- IGBP Land Cover Classification (Belward and Loveland, 1995)
- U.S. Geological Survey Land Use/Land Cover System (Anderson and others, 1976)
- Simple Biosphere Model (Sellers and others, 1986)
- Biosphere-Atmosphere Transfer Scheme (Dickinson and others, 1986)

For this database, Olson Global Ecosystems Legend system was chosen for its moderate details of the annual land cover characterization.

The file can be directly downloaded from the web site http://edcwww.cr.usgs.gov/landdaac/glcc/tablambert_na.html or ftp the file from the web site ftp://edcftp.cr.usgs.gov/pub/data/glcc/na/lambert/. The file name is naoge1_01l.img.gz. The following section provides the detailed steps for processing the GLCC data. More detailed information may also be found on the data description web site http://edcwww.cr.usgs.gov/landdaac/glcc/nadoc.html.

1. Unzip the file using gunzip in UNIX.

2. Rename the file name naoge1_01l.img to naogel_01l.bil

3. Create the header file: naogel_o1l.hdr (text file) and put it in the same directory as file naogel_01l.bil

```
nrows 8996
ncols 9223
nbits 8
ulxmap -4486549.963
ulymap 4479868.288
xdim 1000
ydim 1000
```

4. Convert the image to grid (naogel) using Arc command image grid

```
Arc: imagegrid naoge1_01l.bil naogel
Converting Image to Grid ...
```

5. Define the projection of the grid

```
Arc: projectdefine grid naogel
Define Projection
Project: projection lambert_azimuth
Project: parameters
radius of the sphere of reference [0.00000 ]: 6370997.00000
longitude of center of projection [  0.000 ]: -100 0 0.0
latitude of center of projection  [  0.000 ]: 50 00 0.0
false easting (meters)            [0.00000 ]: 0.0
false northing (meters)           [0.00000 ]: 0.0
```

6. Retrieve the regional data of interest.

   The naogel grid contains data for all North America. The data for the region of interest can be retrieved using Grid command: setwindow.

```
Arc: grid
Grid: display 9999
Grid: mape naogel
Grid: gridpaint naogel
Grid: setwindow *
< Use the curser to draw the region of interest>
```

```
Grid: mygrid = naogel
Grid: clear
```

7. Convert the grid to polygon coverage

   The vegetation coverage from GLCC can be converted from grid using Arc command *gridpoly*.

```
Grid: gridpoly mygrid mycover
```

8. Clip out the vegetation R_vege1k for the study region using template cover *cuttersp*.

9. Edit the attribute table of vegetation theme in ArcView to add in the attribute information about the vegetation types according to the Olson Global Ecosystems. The value description of the Olson Global Ecosystems is provided at the Appendix D. It is also provided on the Web Site:

   (http://edcwww.cr.usgs.gov/landdaac/glcc/nadoc1_2.html#olson)

Since the vegetation data provided by GLCC are based on 1- km Advanced Very High Resolution Radiometer data spanning, the accuracy of the data needs to be justified by comparing the data with the actual vegetation distribution.

The vegetation for the study region from the GLCC is shown in Figure 4.11.

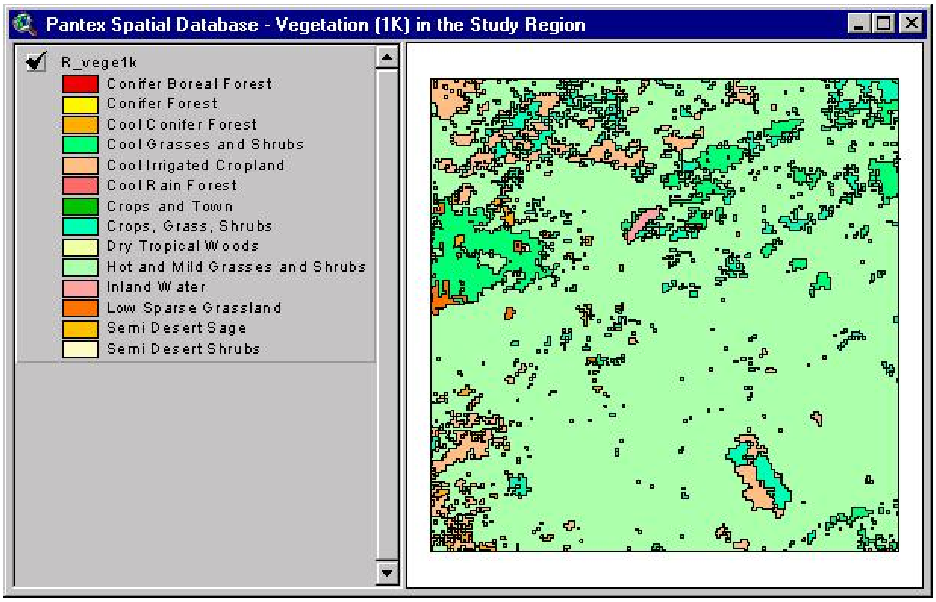


Figure 4.11 Vegetation Description of the Study Region Provided by GLCC

4.6.2 Vegetation Data from TNRIS

The vegetation data downloaded from TNRIS Vegetation (.e00 format) was the vegetation coverage for the State of Texas obtained from Texas Parks and Wildlife. The Texas State Mapping System coordinate system description of the coverage is provided below:

COORDINATE SYSTEM DESCRIPTION

```
Projection          LAMBERT
Units               METERS     Spheroid         GRS1980
Parameters:
1st standard parallel              34 55  0.000
2nd standard parallel              27 25  0.000
central meridian                 -100  0  0.00
latitude of projection's origin    31 10  0.000
false easting (meters)              1000000.00000
false northing (meters)             1000000.00000
```

The compressed e00 file vegcov.e00.gz was downloaded from the TNRIS ftp site Vegetation Types: ftp://www.tnris.state.tx.us/pub/GIS/vegetation/. The file was uncompressed using gunzip in UNIX and imported to ARC/INFO as vegetation coverage. Then the coverage was projected to State Plane Coordinates and the Pantex regional vegetation coverage was clipped from the projected coverage using the template *uttersp* for the study region. The vegetation for the study region from the TNRIS vegetation data is shown in Figure 4.12.

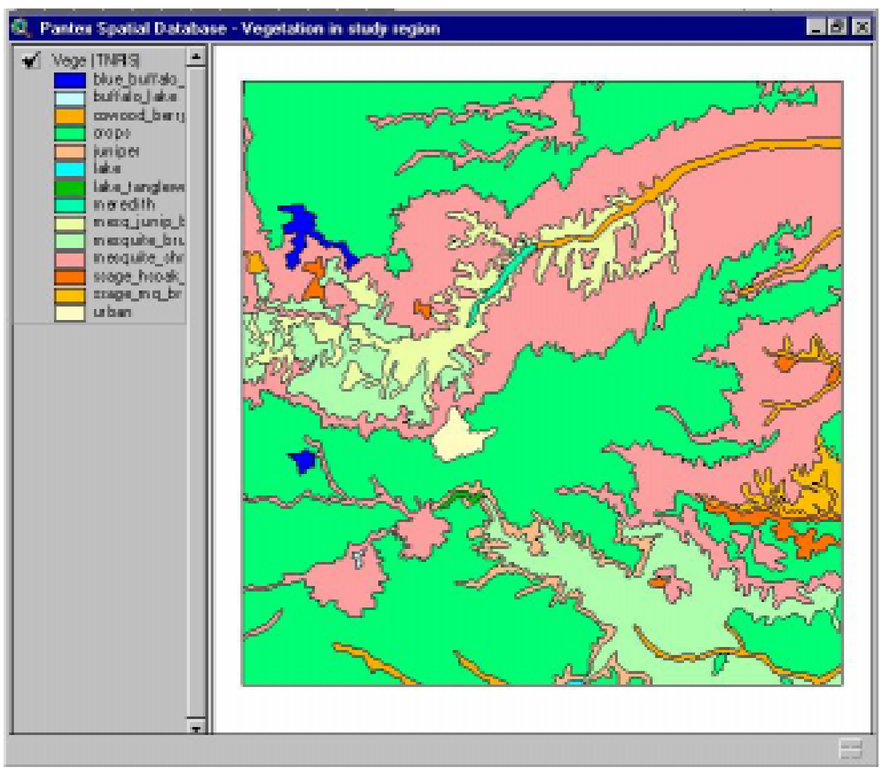


Figure 4.12 Vegetation Description of the Study Region Provided by TNRIS

4.7 AQUIFER DATA

Aquifer data were also obtained from TNRIS ftp site, Aquifers, Major and Minor (e00format). The data provided are presented in the individual coverage for each aquifer. Since these coverages are overlaid with each other, a continuous coverage cannot be constructed for the study region. In this spatial database, the aquifers that cover or partially cover the study region were identified, each of them was projected into the State Plane Coordinates. The part of the aquifer that lies within the study region was then clipped using the template coverage of study region *cuttersp*. There are two aquifers (Ogallala and Dockum) that have overlap with the study region. The aquifer data in the study region are shown in Figure 4.13.

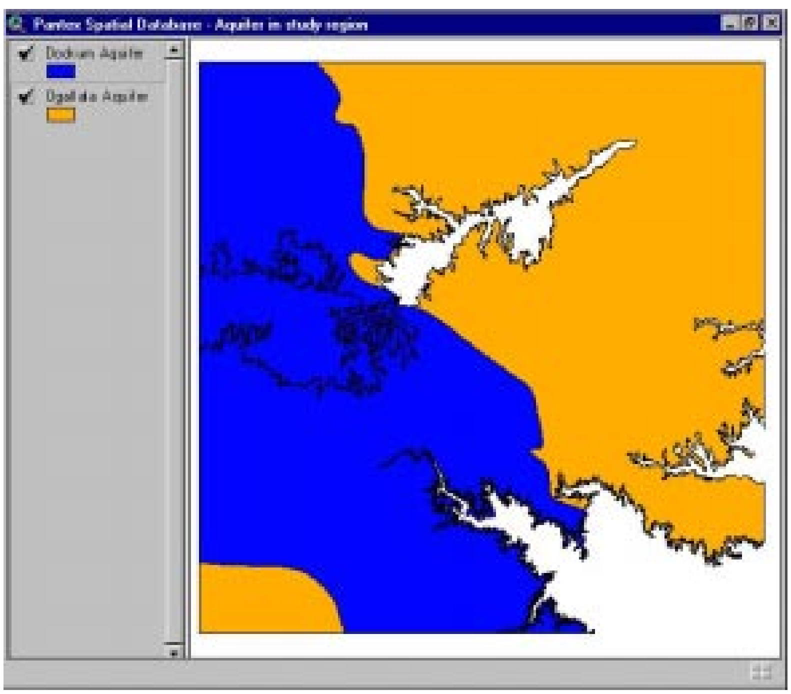


Figure 4.13 Aquifer Data in the Study Region

Extensive groundwater monitoring work has been conducted in Pantex facility and the surrounding area. These data can also be incorporated into the aquifer database. However, currently these data has not been transferred over from the Pantex facility. This part of the database will need to be improved in the future.

4.8 PANTEX FACILITY DATA (CAD)

The Pantex facility CAD converted from the facility CAD file were provided by Pantex GIS specialist GARY L. THOMAS GTHOMAS@pantex.com.

There are five different types of data in the Pantex Facility Data Sets:

- buildings.dgn (plant map)
- 5foot.dgn (5 ft contour map)
- contours.dgn (1 foot contour map)
- 7mileba.dgn (USGS Quad Map)
- 8x11base (plant outline used for low detail)

The most important data set is the building layout (buildings.dgn) which is shown in Figure 4.14, however, attribute information is not provided in the data set, which makes it difficult to interpret the layout. The other data sets are shown in Figure 4.15 to 4.18.

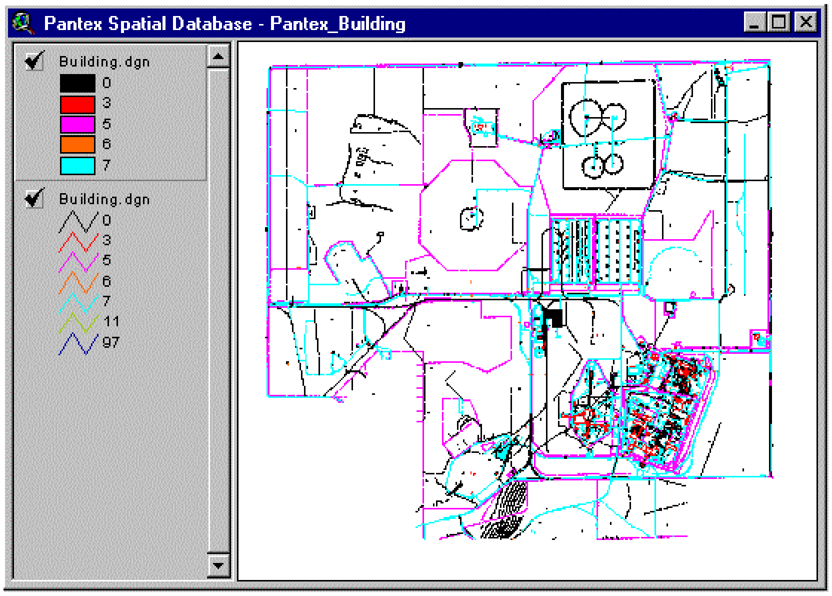


Figure 4.14 Pantex Building CAD Map in ArcView

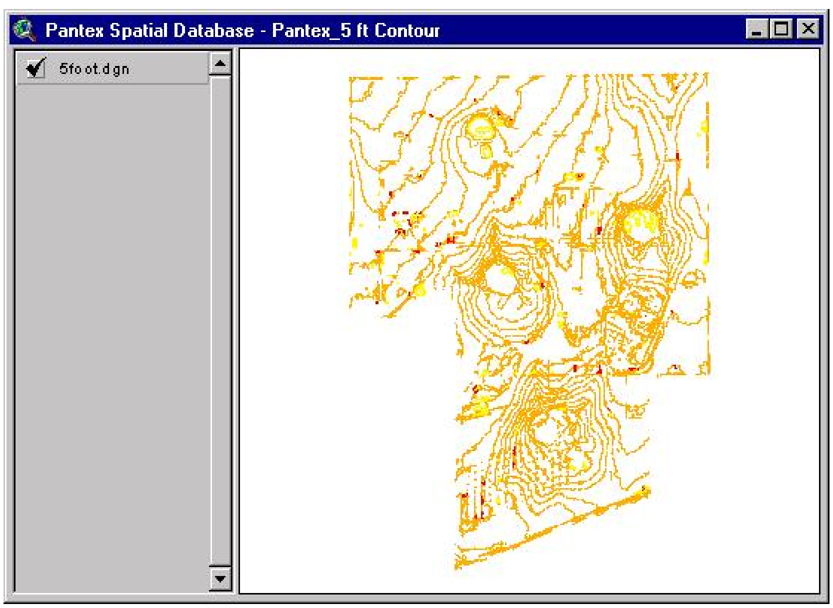


Figure 4.15 Pantex 5-ft Contour Map

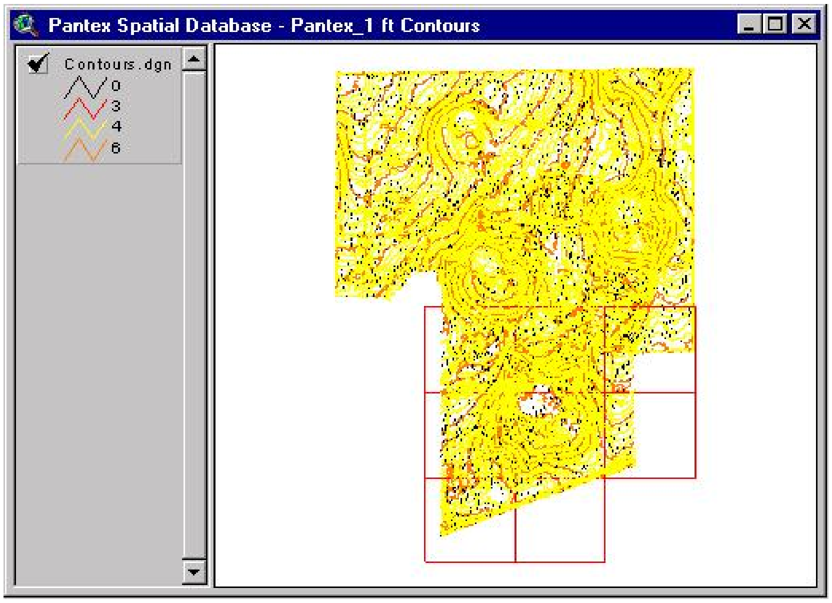


Figure 4.16 Pantex 1-ft Contour Map

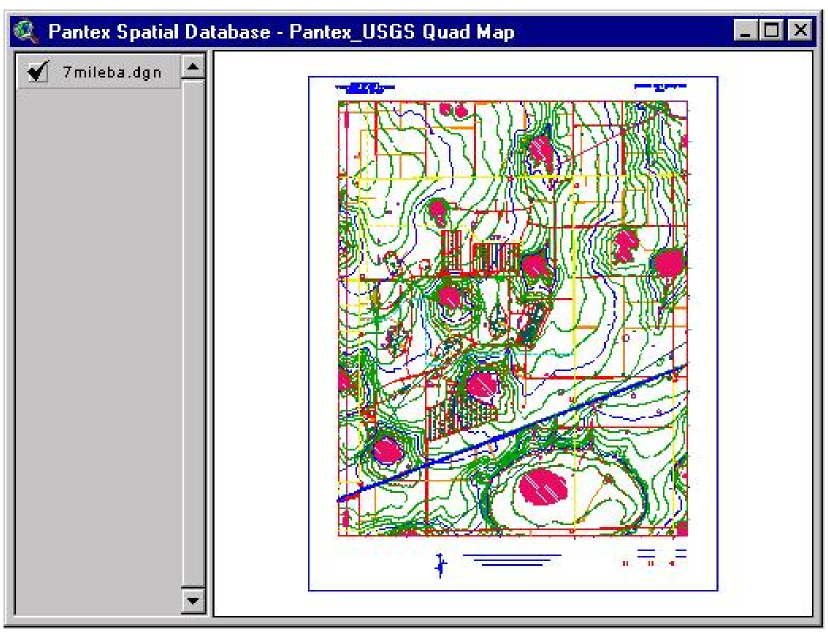


Figure 4.17 Pantex USGS Quadrangle Map

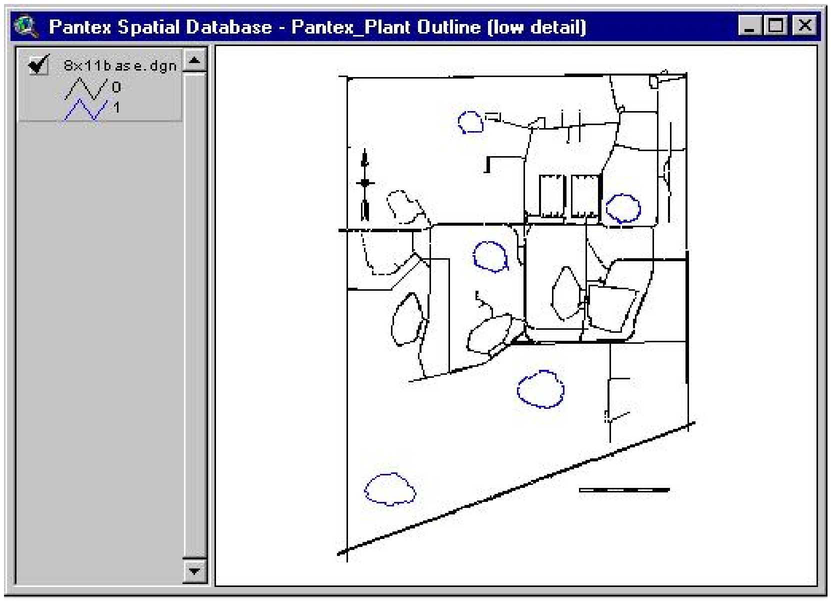


Figure 4.18 Pantex Plant Outline (Low Detail)

In summary, the Pantex spatial database was designed to provide the spatial database framework for characterization of environmental risks from the MOX fuel processing facility at the Pantex Plant. It is also designed to fill information gaps so that additional researches may be initiated in a timely fashion, and informed decisions may ultimately be made. The database currently contains eight data layers pertaining to environmental projects. (Figure 4.19) The database is in a preliminary condition, with limited data layers, a simple work directory database structure, and no environmental operating and sampling data included. The data dictionary is provided at Appendix E. It is also provided at the web site: http://www.ce.utexas.edu/prof/maidment/intranet/pantex/pubwin/CD-ROMdoc.html.

Figure 4.19 Eight Data Layers in the Pantex Spatial Database

Current the database is provided in a CD-ROM format. The spatial database has a total volume of less than 300 MB. A CD-ROM is sufficient enough to store all of the data. The data are presented in spatial data file format (e.g., coverage, grid, image and shapefile). In addition to the advantages mentioned previously of using CD-ROMs as a database distribution alternative, the CD-ROMs can also be used as a spatial database management system for a relatively smaller spatial database. With this approach, a customized spatial database with a simpler database structure is possible for a specific project. Since a large relational database (e.g., Oracle and Sybase) is expensive to implement and maintain, Microsoft Access can be used as an external relational database for a CD-ROM spatial database. This approach will be presented in the Chapter 5.

Future developments of the Pantex spatial database will involve developing more data sets (DEM and watershed, precipitation, wind speed, etc.), reconstructing the database structure (Access, ArcStorm, Oracle with SDE, or others), and incorporating the environmental operating and sampling data into the database.

Chapter 5: ADDITIONAL DATABASE COMPONENTS AND TECHNOLOGIES

5.1 INTEGRATING MICROSOFT ACCESS WITH ARCVIEW

5.1.1 Introduction

As mentioned in the previous section, environmental sampling data needs to be incorporated into the spatial database. We need to consider three intrinsic features of the sampling data when we try to decide the database structure for them.

First, these sampling data have the sampling *locations* associated, which are presented as latitude and longitude. Therefore these data can be stored in a spatial database. The easiest way to present the sampling locations is to generate a point coverage using the latitude and longitude of the sampling sites. Secondly, there may be many *attributes* associated with each of the sampling locations, such as different constituent concentrations, which increases the complexity of the database. But this can still be handled by using the multiple fields of the attribute table of the point coverage. Thirdly, the sampling data are usually generated regularly, which means there may be a third dimension – the *temporal* dimension. The third dimension makes the database really complicated. A two-dimensional attribute table in ArcView has some difficulty to handle time as well.

The easiest way to handle these three dimensional or multidimensional data would be to use multiple tables that are linked by a primary key – the relational database. ArcView has the ability to link tables; however, its ability to deal with many tables with large amount of records is limited compared to an external relational database base management system. An external relational database also provides better data management such as data entry, updating, query, backup, recovery and networking capability. If we can make a smooth connection between ArcView and an external relational database, we can store and manage all of the sampling data in the external relational database, select a subset of the data of interest by query, and display and analyze them in ArcView. Therefore we can take the advantage of the data management features of the relational database, and the spatial data display and analysis features of ArcView.

In this project, we proposed to use Microsoft Access as the external relational database management system for environmental sampling data for the following reasons:

1) It is readily available as part of the MS Office, which is provided on most desktop PCs.
2) It is very easy to use with many built in wizards. No specific training is required. Most people can quickly learn how to use it with the aids of the online help or the menus.
3) It is much cheaper than many other databases such as Oracle or Sybase.
4) It can be easily connected to other software in MS office such as MS Word or Excel. It can directly generate reports as word document.
5) It provides macros and programming modules (Visual Basic) to automate database operations.

5.1.2 Objectives

The objective of this project is to integrate the Microsoft Access to ArcView, so we can store all the environmental sampling data in Access and use ArcView to display and analyze a subset of the data of interest.

5.1.3 Methodology

The schematic diagram of the integration is shown in the Figure 5.1

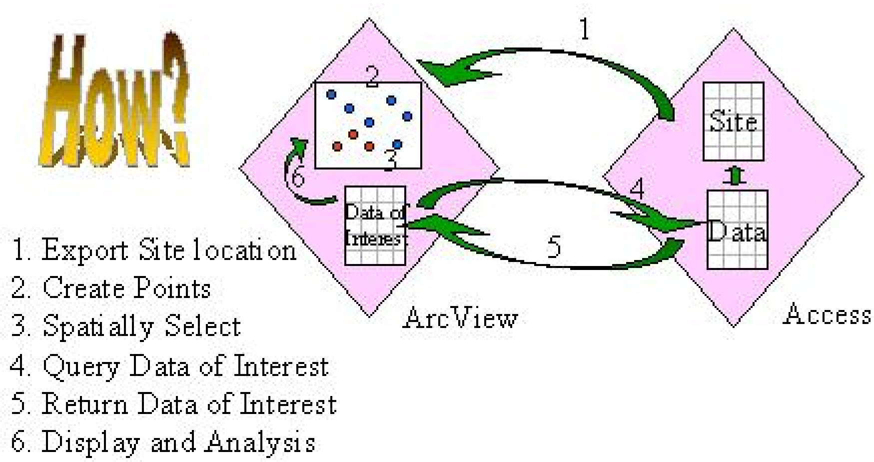


Figure 5.1 Schematic Diagram of the Integration of Access with ArcView

The environmental sampling data were stored as two tables in Microsoft Access database. The *Site* table stores the spatial information of the sampling sites as latitudes and longitudes. The *Data* table stores the water quality data of different constituents at various sampling times. The sampling site locations were exported to ArcView, and the point coverage of the site locations was created. The locations of interest were spatially selected in ArcView and exported to

Access. Specific queries were performed in Access to retrieve the data of interest. The retrieved information was displayed and analyzed in ArcView.

The Access database includes preconstructed queries and macros and customized menu items to link the ArcView and Access. An ArcView project is also provided with Avenue scripts and customized user interface to help create the site locations as a point coverage and export the selection from ArcView.

5.1.4 Procedure and Results

The following section provides the basic procedure to integrate Microsoft Access with ArcView:

1) Export the sampling site locations in MS Access to a site dBASE table called *site.dbf* using the customized menu in Access database (*AccessAv/Export Site Locations to ArcView*)
2) Add in the site.dbf table to ArcView using Table/Add
3) Create a point coverage *sampsite.shp* of the site locations using the customized menu in ArcView project (*AvAccess/Create Point Coverage*).
4) Selected the locations of interest.
5) Export selection using the customized menu in ArcView project (*AvAssess/Export Selection*)
6) Make sure that there is no *avselect* table in Access database (if yes, delete it), import the selection from ArcView using the customized

menu in the Access database (*AccessAv/Import Selection from ArcView*).

7) Construct queries to retrieve the data of interest for those spatially selected locations. The queries need to be set up at the first time. They can be saved for the future use. Export the query results to ArcView using the customized menu in Access database menu (*AccessAv/Export Query Result to ArcView*).
8) Add the query result table in ArcView and join it to the attribute table of *sampsite.shp*
9) Display the attribute information of the selected locations by Autolabel or identify tool.
10) Perform the analysis on the selection. For example, creating the contour of nitrate concentration in the ground water of the Pantex Facility will result the following view shown in Figure 5.2.

More detailed procedure is provided in the exercise: *Integrating Access into ArcView*, which is attached in the Appendix F. The ArcView and Access files will be included in CD-ROM format of the report.

In ArcView project files, two Avenue scripts were used to generate the point coverage (crtpntdd.ave) and export the selection of site to dBase file (AvAccess.export). These two scripts are provided in Appendix C.

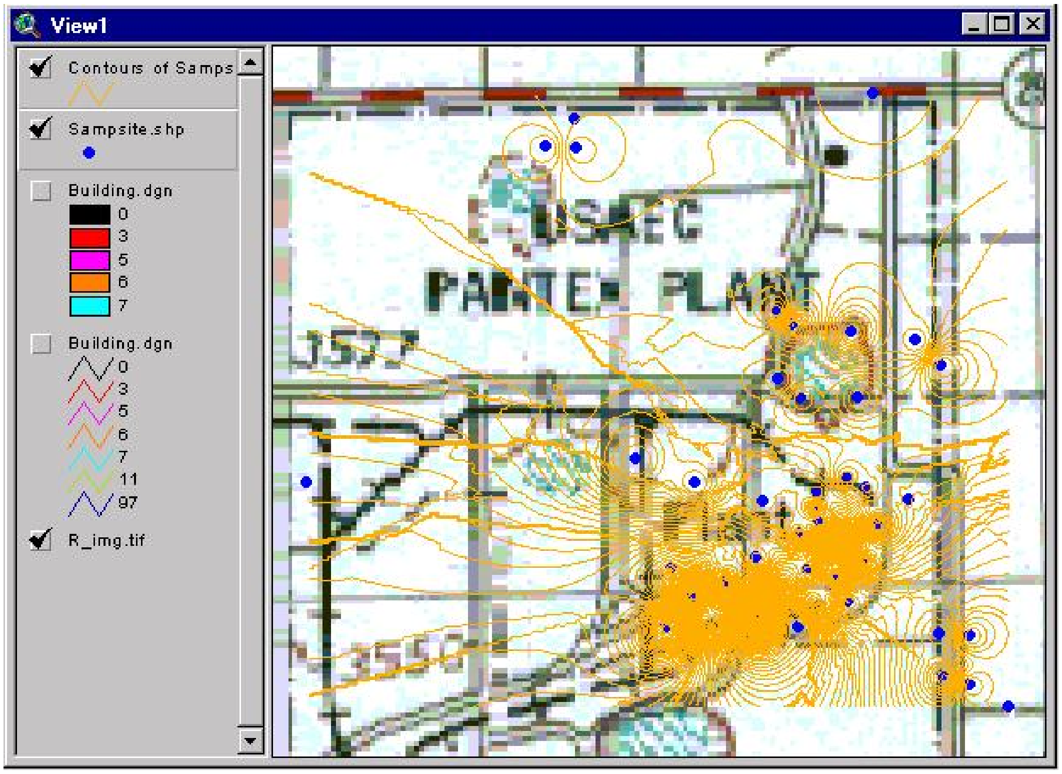


Figure 5.2 Nitrate Concentration Contours in the Groundwater of the Pantex Facility

5.2 INTERNET MAP SERVER

The Internet is the world's largest network. It is open, inexpensive, easy to use, platform-independent, and supports multimedia. User-friendly Internet browsers have made the Internet very popular with the general public. Some of the most obvious potential uses of the Internet for GIS are data publishing, product sales and distribution, GIS services. GIS functionality, as an information system to manage and visualize geographic data, is in great demand across the Internet. Today, ESRI offers extensions to MapObjects, ArcView, and the Spatial Database Engine (SDE) software products to allow for their utilization over the Internet.

The extensions to MapObjects and ArcView are called Internet Map Server (IMS) extensions. Internet Map Server is a powerful technology for delivering the mapping and GIS application over the Internet.

5.2.1 MapObjects Internet Map Server

MapObjects is a collection of GIS objects that conform to Microsoft's Active X OLE/COM specification. MapObjects can be embedded in other applications as specific-purpose mapping and GIS components. Internet Map Server is the system that allows any application created with MapOjbects be deployed over a network. The MapObjects-based application acts as a server delivering mapping and GIS services to Internet clients. A schematic diagram of the MapObjects Internet Map Server is shown in Figure 5.3.

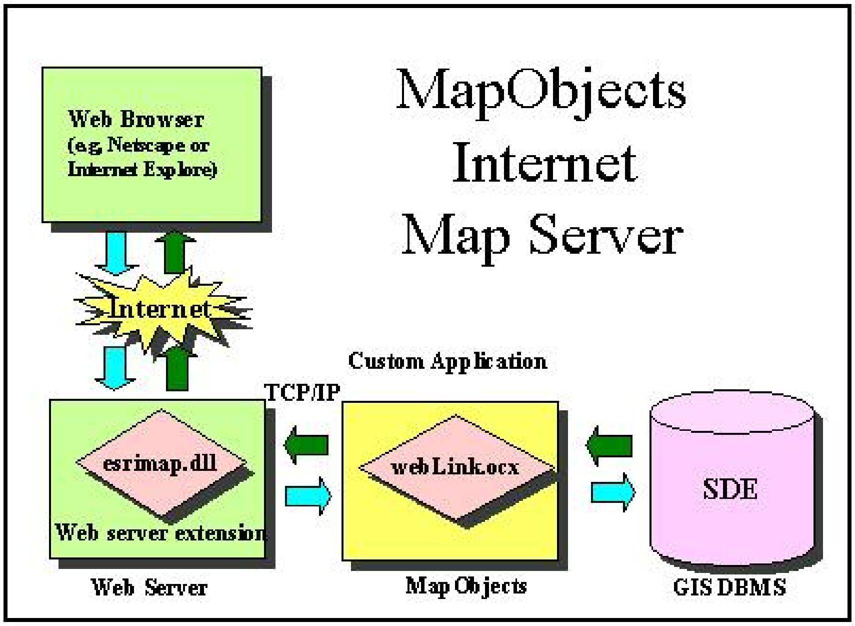


Figure 5.3 MapObjects Internet Map Server (ESRI White Paper Series, 1997b)

The MapObjects Internet Map Server has two basic parts: *esrimap.dll* and *weblink.ocx*. The *esrimap.dll* software allows MapObjects to interface with a standard Internet Web Server. It also provides load-balancing facilities to queue and distribute browser requests equitably where multiple servers are employed. The *weblink.ocx* extension to MapObjects is responsible for copying the applications's display and passing it to the Web server for transmission to a browser.

As Figure 5.3 shows, the users' Web browser communicates with a MapObjects application via a standard Web server. The *esrimap.dll* software enables the standard Web server to communicate with a MapObjects application. Requests from the users are passed to the MapObjects application and processed by the MapObjects application. The results are returned back to the web server though *weblink.ocx* software as raw data, HTML files, or GIS or JPEG images, which can be viewed by standard web browsers.

The MapObjects Internet Map Server is a simple, yet very flexible and powerful technology that becomes extremely popular and widely deployed. Some the MapObjects applications are demonstrated in ESRI MapObjects Internet Map Server live demo web site (http://maps.esri.com/ESRI/mapobjects/demos.htm).

5.2.2 ArcView Internet Map Server

The ArcView Internet Map Server operates in a very similar fashion to the MapObjects Internet Map Servers. ArcView as ESRI's popular desktop GIS software has many extensions, one of which is the Internet Map Server (IMS) extension. The ArcView Internet Map Server extension has an open socket that listens for requests coming from the ESRIMap Web server extension (*esrimap.dll*). The main functional difference between the ArcView Internet Map Server and the MapObjects Internet Map Server is that the ArcView Internet Map Server has an additional piece of software called *MapCafé* to aid user interaction.

Since current browsers and versions of HTML provided very limited ability for user interaction with map data, ESRI has developed a Java applet, called *MapCafé*, which supports client-processing (e.g., zooming, selection of features, attribute queries, hyperlinks to other Web sites, print). *MapCafé* is automatically downloaded into a user's standard browser when the user connects to an ArcView-IMS enabled Web site. This provides an interactive mapping/GIS interface to ArcView GIS, which resides on the Web server or the designated computer for the ArcView applications. The schematic diagram of ArcView IMS is shown in Figure 5.4.

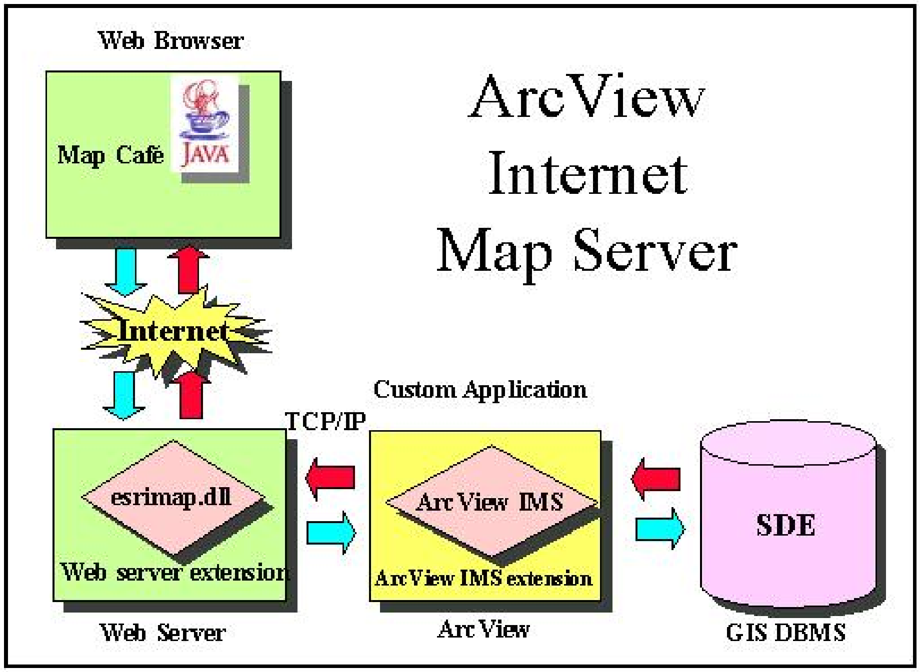


Figure 5.4 ArcView Internet Map Server

Compared to the MapObjects IMS, the ArcView IMS has the advantages of ease of use, out-of-box utility, and multiplatform support. No programming is required for using ArcView IMS. The extension automatically creates a ready-to-use Web page containing the map and an interactive user interface that people can use to browse, explore, and query the map. The built-in wizard lets you choose which buttons and tools appear on the page, the size of the page and layout of the page. The web page can also be customized by adding text, graphics, backgrounds, and links. You can also use Avenue programming language to customize the functions of the buttons and tools on the map. MapCafé is a ready-to-use, yet configurable Java mapping applet. It provides a standard set of mapping tools for users not only to browse any map published by the map server, but also to pan, zoom, identify the attributes, locate sites on maps and hyperlink through the Web. In addition, Java classes are provided allowing Java programmers to change the response of MapCafé. Figure 5.5 shows a typical map web page created and served by ArcView IMS.

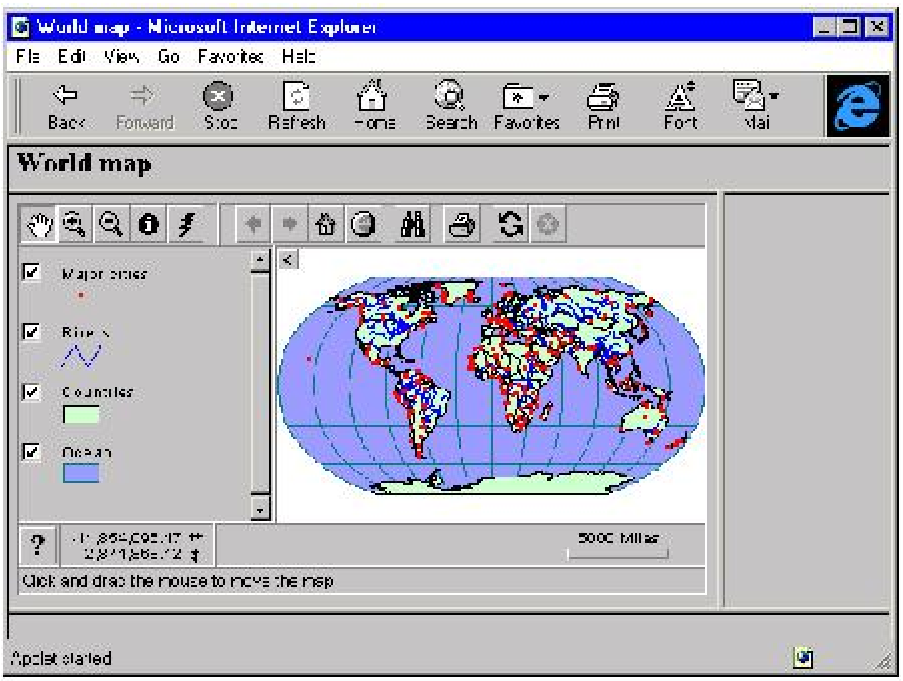


Figure 5.5 Sample web page created by ArcView IMS (ESRI White Paper Series, 1997c)

In order for a user to be able to interactively browse the maps provided by the Internet Map Server, the ArcView Application has to be running. There are several ways to configure the ArcView Application and the web server. For very light server loads, the application can be run on the Web server computer. Otherwise, the application can be run on a separate computer which is connected to the Web server. For heavily requested mapping application, several computers that have ArcView applications running can be connected to the Web server. The Web server automatically balances the load between them, therefore the speed and capacities will be maximized.

5.2.3 Comparison of the MapObjects and ArcView Internet Map Servers

The ArcView Internet Map Server is an easy to use and out-of-box solution for Internet GIS. No programming skill is required. It also provides a certain level of customization. However, it has a heavy overhead and is relatively expensive, computationally. On the other hand, the MapObjects Internet Map Server is a GIS programming tool, which is not for end users but rather for developers. The advantages of the MapObjects Internet Map Server are light overhead, less running CPU time required, Active X compatibility, easy to deploy and relatively cheaper compared to the ArcView Internet Map Server. For Internet mapping applications that may be heavily requested by web browsers but require relatively simpler GIS functionality, the MapObjects Internet Map Server can be an ideal choice. Users requests from the Internet are handled in the order by the mapping applications reside on the server or the application computers. If all of the copies of mapping applications are busy, a user request will be put on hold until the next available mapping application. Since more copies of MapObjects mapping applications can be run on the same computer than that of ArcView map applications, the response to the users' requests will be faster with MapObjects Internet Map Server.

The Table 5.1 indicates the differences between the MapObjects Internet Map Server and ArcView Internet Map Server.

Table 5.1 Differences between ArcView IMS and MapObjects IMS

Features		**ArcView IMS**	**MapObject IMS**
Ease of Use		Easy (Built-in Wizard)	Moderate
Programming Requirement		No	Yes
Supported Platforms	**Server side**	UNIX and PC	PC
	Client side	UNIX and PC	UNIX and PC
Running CPU Time		More	Less
Cost		Relatively expensive	Relatively cheaper

5.3 PANTEX MOX FUEL FACILITY RISK ASSESSMENT RESEARCH INTRANET

Since the end of the cold war, the U.S. and Russia dismantle more and more of their nuclear weapons. But as both nations take apart bombs, plutonium from these devices must be disposed of in such a way that it can never be re-used in nuclear weapons.

The president of the Unite State has declared 38.2 metric tons of this weapons grade plutonium to be "surplus," meaning that the plutonium will never again be used in weapons. More than 21 tons of this surplus plutonium are stored at Pantex.

The Department of Energy (DOE) decided in January 1997 that some of the surplus plutonium will be immobilized (encased in highly radioactive glass or ceramic and buried) and some will be fabricated into mixed-oxide fuel to provide power for electricity generating nuclear power plants while, for all practical purposes, making the plutonium unusable for weapons. Pantex is a possible location for mixed-oxide fabrication because of the large amount of weapons plutonium that is already stored on site. DOE will select the site(s) to fabricate the mixed-oxide only after extensive environmental review, weighing community views and levels of support for taking on the job.

A risk assessment of Plutonium MOX fuel facility at Pantex has been conducted to as part of the information to help DOE to make the decision. This project is a joint effort that involve the experts from Amarillo National Plutonium Research Center (ANRCP), Center for Research in Water Resource (CRWR), Agricultural Research & Extension Center in Texas A&M system, Los Alamos

National Research Lab and the Pantex Plant. Therefore the coordination of the activities and sharing of the information are extremely important. Since the final result of the risk assessment will be presented to the public through the Internet, it would be helpful these results would be previewed internally before they are exposed to the public.

A research Intranet was set up to serve as the following purposes:

- Define the objective and scope of the project
- Clarify the individual function of each team in the project
- Set up a channel for easy communication
- Provide the online resources for each technical subject
- Update the progress report
- Preview the public information

The Intranet provides the essential information about the project such as the project statement, the teams' information, the technical subjects, and the progress reports, it also provides a discussion forum where people involved in the project can directly post necessary information. Most importantly, the Intranet also contains a public window from which the final result of the risk assessment and the spatial database can be presented to the public. Currently the spatial database web site contains *mapbook* (map collections for the Pantex facility) and *CD-ROM* (the documentation of the CD-ROM and sample image of the data set in the spatial database). It also has a link to the Internet Map Server, which will be set up in the future.

The Intranet can be viewed at the web site (currently it is password protected) http://www.ce.utexas.edu/prof/maidment/intranet/pantex/intro.html.

The original Intranet pages can also be viewed locally in the CD-ROM format of this report.

Chapter 6: CONCLUSION

Pantex is America's only nuclear weapon assembly and disassembly site. Environmental impact, environmental safety and public concern are some of the most important issues for the Pantex facility. An environmental information system framework was conceptually designed to help study those issues. Each of the components of the framework is discussed in detail. Several preliminary investigations on some of the components were also conducted. These investigations include 1) developing a spatial database that covers the Pantex facility and the surrounding region and distributing the database in CD-ROM format, 2) integrating MS Access as an external relational database to ArcView, 3) investigating the feasibility of implementing an Internet Map Server, and 4) setting up an research intranet for Risk Assessment of Pantex Plutonium MOX Fuel Facility.

From the investigations of the components of Environmental Information System, several conclusions were formed and are listed below:

1) A GIS-based environmental information system can be very beneficial for the Pantex facility or a similar type of facility to compile and organize the environmental information so that information gaps will be filled, information resources will be conserved and environmental projects can be conducted in a timely manner.

2) As computer and network technology matures, more and more information is readily available on the Internet. Building a spatial database using information directly downloaded from the Internet is feasible, efficient and very cost effective.
3) The CD-ROM approach as a data distribution alternative for the spatial database server and the database management alternative for the individual user is effective. The major advantages of this approach are to relieve the burden on the server and to provide the users with customized and simpler databases to meet individual needs.
4) Microsoft Access as an easy-to-use, yet very powerful Relational Database Management System can be directly integrated with ArcView to produce an Environmental Information System on a PC with moderate storage requirements. The spatial and attribute information can be stored and managed in Access in a tabular format, while a subset of data with the information of interest can be displayed and analyzed in ArcView. Therefore, the advantages of database management of Access and the advantages of spatial display and analysis of ArcView can be combined.
5) Internet Map Server is a powerful technology that can deliver GIS functionality over the Internet. The two currently available options MapObjects IMS and ArcView IMS provide functionality that is complementary to one another. Users can make their choice based on the complexity and flexibility of the projects, the frequency of the

browsers' requests, the programming level of the developer, and the budget of the project.

6) The Internet as the world's largest network not only provides the vast amount of information, but can also serve as a research communication and training channel.

The investigation of the Environmental Information System of Pantex is incomplete. Future work will involve the following tasks:

1) Improve the current Pantex spatial database by incorporating other types of regional description data such as a digital elevation model and wind speed data.
2) Incorporate the environmental sampling data as Access Tables into the spatial database.
3) Improve the connection between Access and ArcView.
4) Set up an Internet Map Server to present the spatial database to the general public
5) Develop a MapObjects application to present the Risk Assessment results.
6) Refine the Intranet and the public window before presenting the Environmental Information System to the public.

References

Pantex, 1997a, *Pantex Historical Perspective*, URL: http://www.pantex.com:80/ds/pxgena1.htm.

Pantex, 1997b, Pantex Home Page, URL: http://www.pantex.com:80/ds/index.htm.

Pantex, 1997c, *Environmental Issues*, http://www.pantex.com/ds/pxeisa2b.htm.

Amarillo Demographics, 1997, URL: http://www.amarillo-tx.com/demog.html#population.

Thomas, Gary (GTHOMAS@pantex.com) and Fitch, Rob (rfitch@pantex.com), Pantex Database information (email on October, 9, 1997).

Leibold, A. M., Blaskowski, M. J., and Farnes, R. J., 1997, *Environmental Data Management and GIS*, ESRI 1997 User conference proceeding, San Diego, California, URL: http://www.esri.com/base/common/userconf/proc97/PROC97/TO550/PAP511/P511.HTM.

BOZÓ, P., KARIG, G., RACZ, T., and TULLNER, T., 1997, *The Integrated GIS of the Ministry for Environment and Regional Policy of Hungary*, ESRI 1997User conference proceeding, San Diego, California, URL: http://www.esri.com/base/common/userconf/proc97/PROC97/TO550/PAP504/P504.HTM.

ESRI 1997 user conference proceeding, 1997, *Environmental and Natural Resources subjects,* San Diego, California, URL: http://www.esri.com/base/common/userconf/proc97/PROC97/TOPLVL/TRACK.HTM.

Schaller, J. and Bonazountas, M., 1996, *ENVIROCITY - Public Online Environmental Information Services for European Cities,* ESRI 1996European User Conference Proceedings, London, England, URL: http://www.esri.com/base/common/userconf/europroc96/PAPERS/PN07/PN07F.HTM.

ESRI Educational Services, 1996a, Managing a GIS, p2-99, Environmental Systems Research Institutes, Redlands, CA.

ESRI Educational Services, 1996b, Managing a GIS, p2-51, Environmental Systems Research Institutes, Redlands, CA.

ESRI Educational Services, 1996c, Managing a GIS, p2-57, Environmental Systems Research Institutes, Redlands, CA.

ESRI Educational Services, 1996d, Managing a GIS, p2-83, Environmental Systems Research Institutes, Redlands, CA.

ESRI Educational Services, 1996e, Managing a GIS, p2-112, Environmental Systems Research Institutes, Redlands, CA.

ESRI White Paper Series, 1997a, *Metadata Management in GIS*, Environmental Systems Research Institutes, Redlands, CA.

ESRI White Paper Series, 1997b, *How ESRI Uses the Internet/Intranet To Deliver GIS On-Line*, Environmental Systems Research Institutes, Redlands, CA.

ESRI White paper Series, 1997c, *ArcView Internet Map Server Extension Functional Overview*, Environmental Systems Research Institutes, Redlands, CA.

USGS, 1997, *Land Use and Land Cover Data User Guide*, URL: http://edcwww.cr.usgs.gov/glis/hyper/guide/1_250_lulc#lulc14.

APPENDIX A PROJECTION FILES

A projection file provides a set of parameters to convert the spatial data from the original projection to the preferred projection. The conversion of projections of spatial data can be done in ARC/INFO by using the *project* command whose parameters can be specified by a dialog, but it is more convenient to type out a text file with all the responses that would be needed in the dialog and input that at the time the *project* command is initiated. In addition, using the projection file can keep the conversions consistent. A projection file usually contains an input section for the original projection parameters and an output section for the preferred projection parameters. Each section includes the following items:

- projection - the type of projection (required)
- units - the units of the spatial coordinate system (required)
- datum or spheroid - the earth datum or spheroid used for the data(optional)
- parameters - the spatial parameters of this particular projection (the command parameters is required but there may be no parameters

A projection file always has *end* as its last line.

Table A-1 provides a list of the projection files used in the Pantex spatial database development. The actual projection files are also attached in the following sections.

Table A-1 The Projection Files Used in the Pantex Spatial Database Development

Projection File Name	Description
albspco	Project from Albers Projection to State Plane Coordinates
geoddsp	Project from Geographic Coordinates in decimal degree to State Plane Coordinates
geodssp	Project from Geographic Coordinates in decimal second to State Plane Coordinates
lamalb	Project from Lambert Projection to Albers Projection
Lamazisp	Project from Lambert_Azimuthal to State Plane Coordinates
lamstapl	Project from Lambert Conformal Conic Projection to State Plane Coordinates
soiltxsp	Project the orginal soils data from Texas State Mapping System in Albers to State Plane Coordinates
spgeodd	Project from State Plane Coordinates to Geographic Coordinates in Decimal Degree
spgeods	Project from State Plane Coordinates to Geographic Coordinates in Decimal Second
utm_e	Peoject from UTM Zone 14 to Geographic Coordinates
utm_w	Peoject from UTM Zone 13 to Geographic Coordinates
splam	Project from State Plane Coordinates to Lambert Projection

albspco

```
input
projection albers
units meters
datum nad83
parameters
27 25 00
34 55 00
-100 00 00
31 10 00
1000000.0
1000000.0
output
Projection STATEPLANE
Fipszone 4201
Datum NAD83
Zunits NO
Units FEET
Spheroid GRS1980
Xshift 0.0000000000
Yshift 0.0000000000
Parameters
End
```

geoddsp

```
Input
projection GEOGRAPHIC
units dd
parameters
output
Projection STATEPLANE
Fipszone 4201
Datum NAD83
Zunits NO
Units FEET
Spheroid GRS1980
Xshift 0.0000000000
Yshift 0.0000000000
Parameters
End
```

geodssp

```
Input
projection GEOGRAPHIC
units ds
parameters
output
Projection STATEPLANE
Fipszone 4201
Datum NAD83
Zunits NO
Units FEET
Spheroid GRS1980
Xshift 0.0000000000
Yshift 0.0000000000
Parameters
End
```

lamalb

```
input
projection lambert
units meters
datum nad83
parameters
34 55 00
27 25 00
-100 00 00
31 10 00
1000000.0
1000000.0
output
projection albers
units meters
datum nad83
parameters
27 25 00
34 55 00
-100 00 00
31 10 00
1000000.0
1000000.0
end
```

Lamazisp

```
input
Projection Lambert_Azimuthal
units meters
parameters
6370997
-100 00 00
50 00 00
0.0
0.0
output
Projection STATEPLANE
Fipszone 4201
Datum NAD83
Zunits NO
Units FEET
Spheroid GRS1980
Xshift 0.0000000000
Yshift 0.0000000000
Parameters
End
```

lamstapl

```
input
projection lambert
units meters
datum nad83
parameters
34 55 00
27 25 00
-100 00 00
31 10 00
1000000.0
1000000.0
output
Projection STATEPLANE
Fipszone 4201
Datum NAD83
Zunits NO
Units FEET
Spheroid GRS1980
Xshift 0.0000000000
Yshift 0.0000000000
Parameters
End
```

soiltxsp

```
input
projection albers
units meters
datum nad27
parameters
29 30 00
45 30 00
-96 00 00
23 00 00
0000000.0
0000000.0
output
Projection STATEPLANE
Fipszone 4201
Datum NAD83
Zunits NO
Units FEET
Spheroid GRS1980
Xshift 0.0000000000
Yshift 0.0000000000
Parameters
End
```

spgeodd

```
Input
Projection STATEPLANE
Fipszone 4201
Datum NAD83
Zunits NO
Units FEET
Spheroid GRS1980
Xshift 0.0000000000
Yshift 0.0000000000
Parameters
output
projection GEOGRAPHIC
units dd
parameters
End
```

spgeods

```
Input
Projection STATEPLANE
Fipszone 4201
Datum NAD83
Zunits NO
Units FEET
Spheroid GRS1980
Xshift 0.0000000000
Yshift 0.0000000000
Parameters
output
projection GEOGRAPHIC
units ds
parameters
End
```

utm_e

```
Input
projection utm
units meters
zone 14
spheroid clarke1866
parameters
output
projection GEOGRAPHIC
units dd
parameters
End
```

utm_w

```
Input
projection utm
units meters
zone 13
spheroid clarke1866
parameters
output
projection GEOGRAPHIC
units dd
parameters
End
```

splam

```
input
Projection STATEPLANE
Fipszone 4201
Datum NAD83
Zunits NO
Units FEET
Spheroid GRS1980
Xshift 0.0000000000
Yshift 0.0000000000
Parameters
output
projection lambert
units meters
datum nad83
parameters
34 55 00
27 25 00
-100 00 00
31 10 00
1000000.0
1000000.0
End
```

APPENDIX B AML SCRIPT

AML Script for merge dlg_hydro in zone 14

```
/* hydmerge.aml

/* Modified version of dlgmerge for merging dlg_hydro in UTM Zone
/*14

/* An ARC AML FOR PREPARING DLG DATA FOR REGIONAL ANALYSIS
/*
/* prepared by Bill Saunders, modified by Ye Maggie Ruan,
/* University of Texas at Austin Center for Research in Water
/* Resources, GIS in Water Resources Research group
/*
/* AML NAME: hydmerge.aml (run from the "Arc" prompt)
/* FUNCTION: Prepares selected DLG data for analysis with respect
/* to a particular hydrologic or political region.
/* INPUTS:
/* -all compressed ("zipped") DLG files corresponding to the
/* region of  interest. These zipped files are downloaded fromthe
/* USGS EROS Data Center at http://sun1.cr.usgs.gov/eros-
/* home.html. Alternatively the DLG files can be accessed from US
/* Geodata 1:100,000-Scale DLG Data Compact Disc (USGS, 1993).
/* -a projection file that will allow for conversion from utm map
/* coordinates to whatever projection is desired.
/* -a polygon coverage delineating the boundary of the hydrologic
/* or political region of interest.
/*
/*****************************************************************
****
/* BEGIN AML EXECUTION
/*
/* Assuming that zipped DLG files have been downloaded from CD-
/* ROM (in this case, the first three characters of the 8 hydro
/* files are: am1, am2, am3, am4, pv1, pv2, pv3, pv4.
/*
/* The first set of commands below MUST ALWAYS BE CHANGED by the
/* user of the AML. Store the number of zipped DLG files into the
/* variable dlgnum. Then, for each zipped DLG file, define
/* sequential variables called dlg# as the first 3 characters of
/* each of the zipped files. Store the name of your projection
/* file (in this case, utmgeo_e) into the variable prjfname.
```

```
/* Store the name of your hydrologic or political boundary
/* coverage (in this case, sanbord) into the variable border.
/* Finally, specify the type of files that you are using -- the
/* only valid entries for this variable (filetype) are hydro,
/* roads, rail, and mtran.

 &sv dlgnum = 8
 &sv dlg1 = am1
 &sv dlg2 = am2
 &sv dlg3 = am3
 &sv dlg4 = am4
 &sv dlg5 = pv1
 &sv dlg6 = pv2
 &sv dlg7 = py3
 &sv dlg8 = py4

 &sv prjfname = /export/database3/pantex/prjfile/utmgeo_e
/* &sv border = /export/database3/pantex/cutter/cuttersp
 &sv filetype = hydro
/*
/*
 &if %filetype% eq hydro &then
 &sv abbr = hy
 &if %filetype% eq roads &then
 &sv abbr = rd
 &if %filetype% eq rail &then
 &sv abbr = rr
 &if %filetype% eq mtran &then
 &sv abbr = mt
/*
/* This part of the AML unzips all of the compressed files to
/* create 15-minute map files. Each 15-minute map file is first
/* converted into an ARC/INFO line coverage. Then, the borders of
/* each of the 15-minute map files are trimmed away from the
/* coverage so that those 15-minute meridians and parallels will
/* not appear in the final appended coverage.

 &sv count = 1
 &do &while %count% le %dlgnum%
 &sv filename = [value dlg%count%]
 &sv count = %count% + 1
 &sys unzip %filename%%filetype%
 &sv count2 = 1
 &do &while %count2% le 8
 &do &while [exists %filename%%abbr%f0%count2% -file]
 dos2unix %filename%%abbr%f0%count2% %filename%%abbr%f0%count2%
 dlgarc optional %filename%%abbr%f0%count2% %filename%f0%count2%
 &sv x = [delete %filename%%abbr%f0%count2% -file]
 build %filename%f0%count2% line
```

```
 reselect %filename%f0%count2% %filename%0%count2% line # line
 res rpoly# > 1
 ~
 n
 y
 res lpoly# > 1
 ~
 n
 n
 kill %filename%f0%count2% all
 &end
 &sv count2 = %count2% + 1
 &end
 &end
/*
/* This part of the AML merges, or "appends", all of the 15-
/* minute map file coverages together and then builds line
/* topology for the resultant coverage, called "bigmap".
/*
 append bigmap
 &sv count = 1
 &do &while %count% le %dlgnum%
 &sv filename = [value dlg%count%]
 &sv count = %count% + 1
 &sv count2 = 1
 &do &while %count2% le 8
 &do &while [exists %filename%0%count2% -cover]
  %filename%0%count2%
 &sv count2 = %count2% + 1
 &end
 &sv count2 = %count2% + 1
 &end
 &end
 ~
 y
 y
 build bigmap line
/*
/* Once "bigmap" has been created, each of the coverages that
/* were merged to build it are no longer necessary. This part of
/* the AML kills off all of the intermediate level coverages used
/* to append "bigmap".
/*
 &sv count = 1
 &do &while %count% le %dlgnum%
 &sv filename = [value dlg%count%]
 &sv count = %count% + 1
 &sv count2 = 1
 &do &while %count2% le 8
```

```
 &do &while [exists %filename%0%count2% -cover]
 kill %filename%0%count2% all
 &sv count2 = %count2% + 1
 &end
 &sv count2 = %count2% + 1
 &end
 &end
/*
/* The "bigmap" coverage is then reprojected to the desired map
/* projection and coordinates. The projection file must be
/* located in the same directory as the coverage being projected.
/*
 project cover bigmap bigprj %prjfname%
/*
/* Finally, a polgyon coverage of the hydrologic or political
/* boundary of interest is used to "clip" out the hydrologic
/* features specific to that region. The final coverage is called
/* "dlgcov".
/*
/* clip bigprj %border% dlgcov line
/* kill bigmap all
/* kill bigprj all
/*
 &return



/********************end of AML****************************
```

APPENDIX C AVENUE SCRIPTS

Crtpnt.ave

```
'----------------------------------------------------------------
'----------------------------------------------------------------
' Name: crtpnt.ave  3/20/97
' Headline:
' Self:
' Returns:
' Description: Create a point shape file from locations specified
' in a table.
'
' Topics:
' Search Keys:
' Requires:
' History: Create by Seann Reed, Modified by Ye Maggie Ruan
'
'----------------------------------------------------------------
'----------------------------------------------------------------

theProject=av.GetProject
theView=av.GetActiveDoc

theDocs=theProject.GetDocs

tabList=List.Make
for each d in theDocs
  if (d.Is(Table)) then
    tabList.Add(d.getname)
  end
end

'--- IDENTIFY INPUT TABLE
intablename=msgbox.choiceasstring(tabList,"Choose table with
Lat/Lon Values","Table")
if (intablename=nil) then
  exit
end

intable=theproject.finddoc(intablename)
invtab=intable.getvtab
infields=invtab.getfields

'--- IDENTIFY INPUT FIELDS
```

```
latfield=msgbox.choiceasstring(inFields,"Choose the latitude
field.","Latitude")
if (latfield=nil) then
  exit
end

lonfield=msgbox.choiceasstring(inFields,"Choose the longitude
field.","Longitude")
if (lonfield=nil) then
  exit
end

idfield=msgbox.choiceasstring(inFields,"Choose id field.","ID")
if (idfield=nil) then
  exit
end

'latfield=invtab.findfield("lat")
'lonfield=invtab.findfield("lon")
'idfield=invtab.findfield("wsid")


'--- READ AND PROCESS DATA
'--- UNITS OF LATITUDE AND LONGITUDE ARE DDMMSS


OutFileName=FileDialog.Put("outfile".asfilename,"*.shp","Output
Shape File" )
if(OutFileName=Nil)then
  exit
end
OutFileName.SetExtension("shp")
OutFTab=FTab.MakeNew(OutFileName,point)
outTheme=Ftheme.make(outftab)

'CREATE FIELDS FOR THE NEW POINT TABLE
outFields=List.Make
outfields.Add(Field.Make("ID",#field_long,6,0))
outFields.Add(Field.Make("Lat",#field_decimal,8,4))
outFields.Add(Field.Make("Lon",#field_decimal,8,4))
outFieldsc=outFields.DeepClone

outftab.addfields(outfieldsc)
theView.addtheme(outTheme)

if(outFtab.CanEdit)then
  outFtab.SetEditable(true)
else
  msgbox.info("Can't edit the output theme.","Error")
```

```
  exit
end

'IDENTIFY FIELDS FOR WRITING
shpField=outFtab.findfield("shape")
oidfield=outftab.findfield("id")
olatfield=outftab.findfield("lat")
olonfield=outftab.findfield("lon")

for each rec in invtab

  id=invtab.returnvalue(idfield,rec).asnumber
   lat=invtab.returnvalue(latfield,rec)
  'chk=msgbox.yesno(id.asstring,"",true)
'  if (chk=false) then
'    exit
'  end
  if ((id>0) and (lat>0)) then
    lat=invtab.returnvalue(latfield,rec).asstring
    lon=invtab.returnvalue(lonfield,rec).asstring
    latdeg = lat.left(2).asnumber + (lat.middle(2,2).asnumber/60)
+
(lat.right(2).asnumber/3600)
    londeg = lon.left(2).asnumber + (lon.middle(2,2).asnumber/60)
+
(lon.right(2).asnumber/3600)
    newrec=outFtab.AddRecord
    pt=point.make(londeg*(-1),latdeg)
    outFtab.Setvalue(shpField,newrec,pt)
    outftab.setvalue(oidfield,newrec,id)
    outftab.setvalue(olatfield,newrec,latdeg)
    outftab.setvalue(olonfield,newrec,(londeg*(-1)))
  end
end
```

AvAccess.Export

```
'--------------------------------------------------------------
'--------------------------------------------------------------
' Name:AvAccess.export  8/18/97
' Headline:
' Self:
' Returns:
' Description: Export the active table of the selection in
'              Arcview to a .dbf file in the same directory of
'              the project file,so it will be imported or linked
'              into Access.
' Topics:
' Search Keys:
' Requires:  The View should be the active document
' History:   Modified by Ye Maggie Ruan from Table.Export
'            The original code requires that the attribute table
'            of the theme being selected is the active theme,
'            which means after the users have selected the sites,
'            they have to open the attribute table of the theme.
'            This is not very convinient. The modified version
'            works without open the attribute.
'            The modified features are:
'            (1) The View is the active theme
'            (2) Add in the Msgbox asking for only one active
'            theme
'
'              theView = av.GetActiveDoc
'              theThemelst = theView.GetActiveThemes
'              If (theThemelst.count >1) Then
'              MsgBox.Info( "You must have only one active
'              theme.", "")
'              Exit
'              Else
'              theTheme = theView.GetActiveThemes.Get(0)
'              end
'
'              theFTab = theTheme.GetFTab
'--------------------------------------------------------------
'--------------------------------------------------------------
theView = av.GetActiveDoc
theThemelst = theView.GetActiveThemes
If (theThemelst.count >1) Then
  MsgBox.Info( "You must have only one active theme.", "")
Exit
Else
theTheme = theView.GetActiveThemes.Get(0)
end
```

```
theClass = DBASE
theFilter = "*.dbf"
theExt = "dbf"
thedir = fn.getcwd

theFileName=FileDialog.Put("thedir/avselect.dbf".asfilename,
                  "*.dbf","Export Table")
if (theFileName = NIL) then return nil end

theFTab = theTheme.GetFTab
ext     = theFileName.GetExtension
if (theClass.IsSubClassOf(DBASE) and (ext <> theExt)) then
  theFileName.SetExtension(theExt)
end

if (theFTab.GetSelection.Count=0) then
  theFTab.Export(theFileName, theClass, FALSE)
  av.ShowMsg("All records written to"++theFileName.GetBaseName)
else
  theFTab.Export(theFileName, theClass, TRUE)
  av.ShowMsg("Selected records written
to"++theFileName.GetBaseName)
end
```

APPENDIX D OLSON GLOBAL ECOSYSTEMS LEGEND

Value Description

1 Urban

2 Low Sparse Grassland

3 Coniferous Forest

4 Deciduous Conifer Forest

5 Deciduous Broadleaf Forest

6 Evergreen Broadleaf Forests

7 Tall Grasses and Shrubs

8 Bare Desert

9 Upland Tundra

10 Irrigated Grassland

11 Semi Desert

12 Glacier Ice

13 Wooded Wet Swamp

14 Inland Water (see Note 1)

15 Sea Water (see Note 1)

16 Shrub Evergreen

17 Shrub Deciduous

18 Mixed Forest and Field

19 Evergreen Forest and Fields

20 Cool Rain Forest

21 Conifer Boreal Forest

22 Cool Conifer Forest

23 Cool Mixed Forest

24 Mixed Forest

25 Cool Broadleaf Forest

26 Deciduous Broadleaf Forest

27 Conifer Forest

28 Montane Tropical Forests

29 Seasonal Tropical Forest

30 Cool Crops and Towns

31 Crops and Town

32 Dry Tropical Woods

33 Tropical Rainforest

34 Tropical Degraded Forest

35 Corn and Beans Cropland

36 Rice Paddy and Field

37 Hot Irrigated Cropland

38 Cool Irrigated Cropland

39 Cold Irrigated Cropland

40 Cool Grasses and Shrubs

41 Hot and Mild Grasses and Shrubs

42 Cold Grassland

43 Savanna (Woods)

44 Mire, Bog, Fen

45 Marsh Wetland

46 Mediterranean Scrub

47 Dry Woody Scrub

48 Dry Evergreen Woods

49 Volcanic Rock

50 Sand Desert

51 Semi Desert Shrubs

52 Semi Desert Sage

53 Barren Tundra

54 Cool Southern Hemisphere Mixed Forests

55 Cool Fields and Woods

56 Forest and Field

57 Cool Forest and Field

58 Fields and Woody Savanna

59 Succulent and Thorn Scrub

60 Small Leaf Mixed Woods

61 Deciduous and Mixed Boreal Forest

62 Narrow Conifers

63 Wooded Tundra

64 Heath Scrub

65 Coastal Wetland, NW

66 Coastal Wetland, NE

67 Coastal Wetland, SE

68 Coastal Wetland, SW

69 Polar and Alpine Desert

70 Glacier Rock

71 Salt Playas

72 Mangrove

73 Water and Island Fringe

74 Land, Water, and Shore (see Note 1)

75 Land and Water, Rivers (see Note 1)

76 Crop and Water Mixtures

77 Southern Hemisphere Conifers

78 Southern Hemisphere Mixed Forest

79 Wet Sclerophylic Forest

80 Coastline Fringe

81 Beaches and Dunes

82 Sparse Dunes and Ridges

83 Bare Coastal Dunes

84 Residual Dunes and Beaches

85 Compound Coastlines

86 Rocky Cliffs and Slopes

87 Sandy Grassland and Shrubs

88 Bamboo

89 Moist Eucalyptus

90 Rain Green Tropical Forest

91 Woody Savanna

92 Broadleaf Crops

93 Grass Crops

94 Crops, Grass, Shrubs

Note 1: In Version 1.2, all water is mapped to one digital value (14).

APPENDIX E PANTEX SPATIAL DATABASE CD-ROM DOCUMENTATION

Pantex spatial database development was part of the Risk Assessment Project of the MOX Fuel Processing Facility of Pantex. This spatial database includes the spatial data sets for the Pantex facility and the surrounding area. The spatial database is presented in three different formats. All of the actual data sets, which include a number of ARC/INFO coverages, grids, and images are presented in the format of a CD-ROM, which is available upon email (yeruan@mail.utexas.edu or maidment@mail.utexas.edu) requests. Some of the data sets are presented in the maps. An Internet map server will be set up to interactively present the data sets.

The method and the procedure of the development of the Pantex spatial database were documented in Chapter 4 of this report.

This CD-ROM contains regional data sets in seven major categories in a range which consists of a 50 x 50 mile square with the Pantex MOX fuel processing facility at the center. All of the regional data sets are originally retrieved from public domain sources (such as the Internet or CD-ROMs) and have been processed into a consistent format. The CD-ROM also contains the Pantex facility site CAD drawing files, which were provided by the Pantex Facility.

The regional spatial data sets are presented in the format of coverages, grids and images. All of the final data sets are in the State Plane coordinates:

```
Projection STATEPLANE
Fipszone 4201
Datum NAD83
Zunits NO
Units FEET
Spheroid GRS1980
Xshift 0.0000000000
Yshift 0.0000000000
Parameters
```

For most of the data types, the preliminary data sets from which the final regional data were retrieved are also provided in this CD-ROM, usually they contain larger ranges and might be in different projections. An Arcview project is also provided in each data directory to help display the spatial data.

This document is intended to serve as a data dictionary for the Pantex Spatial Database CD-ROM. A description of each file and some sample graphics are also provided. The following naming conventions are used throughout the document:

File Type:

[pC]Point Coverage

[AC] Arc Coverage

[PC] Polygon Coverage

[Gr] Grid

[Img] Image

[CAD] CAD drawing file

[Proj] Arcview Project file

[DBF] DBase Table

[AML] AML file

Projections

(SP) State Plane Projection

(ALB) Albers Equal Area Projection

(UTM) Universal Transverse Mercator Projection

(LAM) Lambert Conformal Conic Projection

(GEO) Geographic Coordinates

In addition to the spatial data, some of the files that were used to help develop or display the spatial database were also included, such as projection files, AML files, legend files, and the template for the regional map. These files can be found in the devkit directory.

CD-ROM Table of Contents

Spatial Database

Aquifer

Census

Digital Line Graph (DLG)

Land use

Pantex Site CAD

Raster Image

Soil

Vegetation

Devkit

Aml

counties

cutter

Legend

Projection

Readme

Data Dictionary

Spatial Database

Directory ***aquifer***

This directory contains aquifer data in Texas panhandle region, two major and two minor. Aquifer Ogallala and Dockum (major) have direct overlap with the study region. Aquifer Blaine and Edtri (minor) are in the surrounding region. The aquifer data were downloaded from TNRIS Anonymous FTP site

The following are the data in the directory aquifer and brief descriptions of the data:

aqdockum [PC] Aquifer Dockum in the study region (SP)

aqogalla [PC] Aquifer Ogallala in the study region (SP)

Aquifer Dockum and Ogallala in the study region

blaine [PC] Aquifer Blaine (LAM)

blainesp [PC] Aquifer Blaine (SP)

dockum [PC] Aquifer Dockum (LAM)

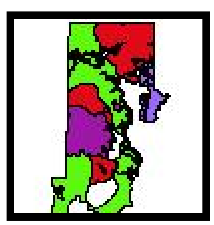
dockumsp [PC] Aquifer Dockum (SP)

edtri [PC] Aquifer Edtri (LAM)

edtrisp [PC] Aquifer Edtri (SP)

ogalla [PC] Aquifer Ogallala (LAM)

ogallasp [PC] Aquifer Ogallala (SP)



Aquifers in the surrounding region

aquifer.apr [Proj] Arcview project file which displays the aquifers in the study region and surrounding area.

Directory ***census***

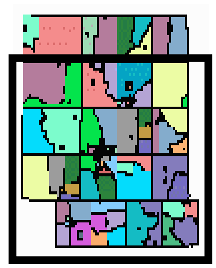
This directory contains census data in the block level of the 20 counties in the Pantex facility surrounding area. The data were developed from the TIGER92 census data that can be download from TNRIS Anonymous FTP site. Census data in the tract level for the each of the twenty counties are also included in this directory.

These data are in Lambert Conformal Conic Projection. The names of the twenty counties are list below:

```
Armstrong
Briscoe
Carson
Castro
Dallam
Deafsmith
Donley
Gray
Hall
Hansford
Hartley
Hutchinson
Moore
Ochiltre
Oldham
Potter
Randall
Roberts
Sherman
Swisher
```

The following are the data in the directory census and brief descriptions of the data.

bigcenlam [PC] Census data in block level for twenty counties around the Pantex facility (LAM) bigcensp [PC] Census data in block level for twenty counties around the Pantex facility (SP)



Census data in block level for twenty counties around the Pantex facility

r_cen [PC] Census data in block level for the study region (SP)

bigcenbk.dbf [DBF] Census data in the block level for 20 counties around the Pantex facility (without spatial attributes data)

bigcensp.dbf [DBF] Census data in the block level for 20 counties around the Pantex facility (with spatial attributes data)

txcensus [DBF] Census data in the block level for whole Texas (without spatial attributes data)

census.apr [Proj] Arcview project file which displays 20 counties census data in block level

Directory ***dlg***

This directory contains 1:100,000 Digital Line Graph (DLG) files which are digital vector representations of USGS 1:100,000-scale, 30- by 60-minute quadrangle maps. The data were downloaded from the CD-ROM with USGS 1:100,000 -Scale file. It can also be downloaded from the USGS ftp site. Twelve

30 by 60-minute quadrangle maps were merged to form the big DLG file from which the regional data were retrieved. Eight of these twelve quadrangles were located in UTM zone 14; the other four were located in UTM zone 13. Each of these twelve quadrangles was merged from eight files using AML file dlgmerge.aml provided in the devkit directory.

Directory ***dlg*** consists of four subdirectories: hydro, roads, rail, mtran. Each has the following meanings:

hydro **–** Hydrography (*Streams and water bodies*)

Roads - Roads and Trails (*e.g., Interstate and primary State highways*)

Rail - Railroads (*e.g., Main line and branch line railroads*)

Mtran -Miscellaneous transportation features (*e.g., Airports and pipelines*)

The following are the data in the directory dlg and brief descriptions of the data.

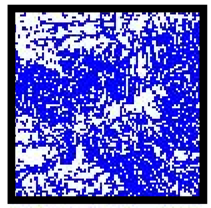
Directory hydro

Bighydgeo [AC] Hydrography coverage which consists of twelve 30 by 60 minutes quadrangle stream coverages (GEO)

Bighydsp [AC] Hydrography coverage which consists of twelve 30 by 60 minutes quadrangle stream coverages (SP)

r_hyd [AC] Hydrography coverage in the study region (SP).

hydmerge.aml [AML] Aml file used to merge the eight hydrography files in UTM zone14 hydmergw.aml [AML] Aml file used to merge the four hydrography files in UTM zone13



Hydrography coverage in the study region

Directory ***roads***

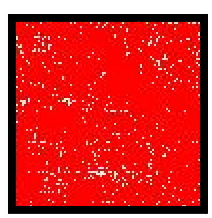
Bigroageo [AC] Roads coverage which consists of twelve 30-by-60 minutes quadrangle roads.(GEO)

Bigroasp [AC] Roads coverage which consists of twelve 30-by-60 minutes quadrangle roads.(SP)

r_roa [AC] Roads coverage in the study region (SP)

roamerge.aml [AML] Aml file used to merge the eight roads files in UTM zone14.

oamergw.aml [AML] Aml file used to merge the four roads files in UTM zone13.



Roads coverage in the study region

Directory ***rail***

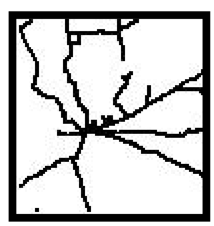
Bigraigeo [AC] Railroads coverage which consists of twelve 30- by -60 minutes quadrangle railroads.(GEO)

Bigraisp [AC] Railroads coverage which consists of twelve 30- by -60 minutes quadrangle railroads.(SP)

r_rai [AC] Railroads coverage in the study region (SP)

raimerge.aml [AML] Aml file used to merge the eight railroads files in UTM zone14.

raimergw.aml [AML] Aml file used to merge the four railroads files in UTM zone13.



Railroad coverage in the study region

Directory ***mtran***

Bigmtrgeo [AC] Railroads coverage which consists of twelve 30- by -60 minutes quadrangle miscellaneous transportation features.(GEO)

Bigmtrsp [AC] Railroads coverage which consists of twelve 30- by -60 minutes quadrangle miscellaneous transportation features.(SP) r_mtr[AC] Miscellaneous transportation features coverage in the study region (SP)

mtrmerge.aml [AML] Aml file used to merge the eight miscellaneous transportation features files in UTM zone14.

mtrmergw.aml [AML] Aml file used to merge the four miscellaneous transportation features files in UTM zone13.

Miscellaneous transportation coverage in the study region

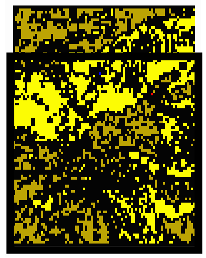
dlg.apr [Proj] Arcview project file which displays all of the dlg files.

Directory ***landuse***

This directory contains land use data, which were compiled from six 1:250k land use coverages downloaded from TNRIS ftp site. The six coverages were merged and the edges were manually corrected in ArcEdit.

The following are the data in the directory landuse and brief descriptions of the data:

biglusp [PC] Land use coverage in Texas panhandle area, which was merged from six (1:250,000) land use files (SP) r_lu [PC] Land use coverage in the study region. (SP)



Land use coverage in the study region

Directory: ***ptxcad***

This directory contains CAD drawings of the Pantex site and facility in the State Plane Coordinates. The CAD files were provided by Pantex facility GIS specialist Gary L Thomas.

The following are the data in the directory ptxcad and brief descriptions of the data.

contours.dgn [CAD] 1 foot contour map of the Pantex site

7mileba.dgn [CAD] Digitized USGS Quad Map

8x11base [CAD] Pantex plant outline used for low detail

buildings.dgn [CAD] Pantex plant map

5foot.dgn [CAD] 5 ft contour map

location.dat, location.ddl and core-dictionary.doc (Microsoft Word for 95) are the database information for sample locations.

Directory ***rasimg***

This directory contains raster image, grids and related files retrieved froma 1:250,000 digital raster image CD-ROM provided by Horizons Technology, Inc..

The following are the data in the directory rasimg and brief descriptions of the data.

inigeo_gr [Gr] Digital Raster Image grid of the surrounding area of the Pantex site. (GEO)

inisp_gr [Gr] Digital Raster Image grid of the surrounding area of the Pantex site from where the regional map was retrieved. (SP)

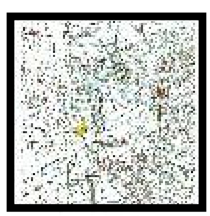
r_img_gr [Gr] Digital Raster Image of the study region. (SP)

color250 The color specification file used in the conversion between grid and image of the raster file.

r_img.tif [IMG] Digital Raster Image for the study region.(SP)

r_img.tfw [IMG] Geo-reference file of the Digital Raster Image.

Rasimg.apr [Proj] Arcview project file, which displays the Digital Raser Image.



Digital Raster Image for the study region.

Directory ***soil***

This directory contains Statsgo soil data in Texas. The soil data were downloaded from TNRIS ftp site.

The following are the data in the directory soil and brief descriptions of the data.

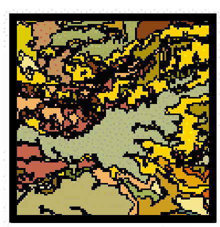
txalb [PC] Statsgo soil coverage for Texas (ALB)

txsp [PC] Statsgo soil coverage for Texas (SP)

r_soil [PC] Statsgo soil coverage for the study region (SP)

soil.apr [Proj] Arcview project file which displays the Statsgo Soil coverages.

In the subdirectory *mapunits*, the information about the soil data (mapunits, components, and layers)are presented as interchangable .e00 files and info tables. For information on how to use the tables, please refer to the web site: http://www.ce.utexas.edu/prof/maidment/CE397/statsgo/viewstat.htm.



Statsgo soil coverage for the study region

Directory ***vegetation***

This directory contains vegetation data from two different sources.

Subdirectory *nadoc* contains vegetation data obtained from North America Land Cover Characteristics Data Base ftp site. Olson Global Ecosystems Legend data set was chosen for the Pantex Spatial database.

Subdirectory *tnris* contains vegetation data obtained from TNRIS ftp site for vegetation data.

The following are the data in the directory vegetation and brief descriptions of the data.

naoge1_01l.hdr Head file used for the conversion from nadoc data set to grid file

vege.apr [Proj] Arcview project which displays vegetation coverages

Subdirectory *nadoc*

bigveg_gr [GR] 1km x1km Vegetation grid of the surrounding area of the Pantex facility.(Lambert_Azimuthal)

bigvegsp [PC] Vegetation coverage from which the regional vegetion data were retrieved (SP)

r_vege1k [PC] 1km x1 km vegetation coverage of the study region (SP)

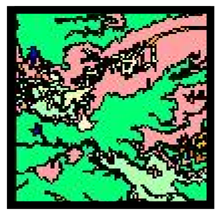
1km x1 km vegetation coverage of the study region

Subdirectory *tnris*

bigveglam [PC] Vegetation coverage of Texas (LAM)

bigvegsp [PC] Vegetation coverage of Texas (SP)

r_vege [PC] Vegetation coverage of the study region [SP]



Vegetation coverage of the study region

Devkit

Directory *Devkit* contains files that were used to help develop the Pantex spatial database, such as AML files, projection files, the template of the study region, the legend files. This directory is provided for the convenience of future development of the Pantex spatial database.

Directory *AML*

dlgmerge.aml [AML] AML files used for merges of the DLG files (upon modification)

wshed.aml [AML] AML files used for watershed delineation (upon modification)

Directory *counties*

ctyln [AC] County line coverage used for identification and calibration of Raster Image

ctylnst[AC] County line coverage used for identification and calibration of Raster Image (SP)

Directory *cutter*

cutgeodd [PC] The template coverage of the study region in decimal degree of geographic coordinates

cuttergeo [PC] The template coverage of the study region in decimal second of geographic coordinates

cuttergr [Gr] The template grid of the study region in decimal second of geographic coordinates

cutterlam [PC] The template coverage of the study region [LAM]

cuttersp [PC] The template coverage of the study region [SP]

Directory *legend*

lgd1_lu.avl: The legend file based on Anderson Land Use Code for the land use coverage.

Directory *prjfile*

This directory contains the projection files used in projection conversions of the Pantex Spatial Database.

albspco: Conversion from [ALB] to [SP]

dlgprj: Conversion from UTM to [ALB] for DLG files

geodssp: Conversion from [GEO] in decimal second to [SP]

geoddsp: Conversion from [GEO] in decimal degree to [SP]

lamalb: Conversion from [LAM] to [ALB]

Lamazisp: Conversion from Lambert_Azimuthal projection to [SP]

Lamstapl: Conversion from [LAM] to [SP]

spgeodd: Conversion from [SP] to [GEO] in decimal degree

spgeods: Conversion from [SP] to [GEO] in decimal second

splam: Conversion from [SP] to [LAM]

utmgeo_e: Conversion from [UTM] zone 14 to [GEO]

utmgeo_w: Conversion from [UTM] zone 13 to [GEO]

APPENDIX F INTEGRATING MICROSOFT ACCESS TO ARCVIEW

Table of Contents

- Goals of the Exercise
- Methodology
- Computer and Data Requirements
- Exercise Procedure

  1. Set File Directory Structure
  2. Build the Environmental Sampling Database
     1. Create the Sampling Data Table
     2. Create the Site Locations Table
  3. Export the Site Locations to ArcView
  4. Create the Point Coverage
  5. Spatially Select the Points of Interest
  6. Import the Selection of ArcView to Access
  7. Query the Data of Interest
  8. Export the Data of Interest to ArcView
  9. Display and Analyze the Data of Interest in ArcView
  10. Employ the integration
  11. Clean up

Goal of the Exercise

This exercise is intended to instruct you on how to integrate the Microsoft Access into ArcView using the Pantex environmental sampling data as an example. First, you will learn how to set up a basic environmental database with two tables: the sampling site table which includes the information about the site locations, and the environment sampling data sets which is associate with these locations. Then you will learn how to set up the link between Access and ArcView, by which you can use Access to store and manage the environmental sampling data, and use ArcView to display and analyze the subset data of interest.

After finishing the exercise, you are expected to understand the basic procedures of creating tables, querys and macros in Microsoft Access. You will also understand how to integrate Access into ArcView.

Methodology

Environmental Sampling data has a complicated data structure. We need to consider three intrinsic features of the sampling data when we try to decide the database structure for it.

First, these sampling data have their sampling locations associated, which are presented as latitude and longitude. Therefore these data can be stored in the spatial database. The easiest way is to generate a point coverage using the latitude and longitude of the sampling site. Secondly, there may be many attributes associated with each of the sampling locations, such as different constituent concentrations, which increases the complexity of the database. But this can still be handled by using the multiple fields of the attribute table of the point coverage or point shapefile. Thirdly, the sampling data are usually generated regularly, which means there may be a third dimension – the temporal dimension. The third dimension makes the database really complicated. A two-dimensional attribute table in ArcView may have some difficulty to handle time as well.

The easiest way to handle these three dimensional or multidimensional data would be to use multiple tables that are linked by the primary key – the relational database. ArcView has the ability to link tables, however, its ability to deal with many tables with a large amount of records is limited compared to an external relational database base management system (RDBMS). An external RDBMS also provides better data management such as data entry, updating, query, backup, recovery and networking capability. If we can make a smooth connection between ArcView and an external RDBMS, we can store and manage all of the sampling data in the external RDBMS, select a subset of the data of interest by query, and display and analyze them in ArcView. Therefore we can take the advantage of the data management features of the RDBMS, and the spatial data display and analysis features of ArcView.

In this exercise, we use Microsoft Access as RDBMS for environmental sampling data for the following reasons:

1. It is readily available as part of the MS Office, which is provided on most desktop PCs.

2. It is very easy to use with many built in wizards. No specific training is required. Most people can quickly learn how to use it aided by the online help or the menus.

3. It is much cheaper than many other databases such as Oracle or Sybase.

4. It can be easily connected to other software in MS office such as MS Word or Excel. It can directly generate reports as Word documents.

5. It provides macros and programming modules (Visual Basic) to automate database operations.

Here is the sample scenario. Initially the environmental sampling data were provided in two basic MS Excel tables. Table 1 stores the spatial information of the sampling sites as latitudes and longitudes. Table 2 stores the water quality data of different constituents at various sampling times. The goal of this project is to store all of the data in Access, to display the sampling locations in ArcView, to select the locations of interest by spatial selection in ArcView, to retrieve the attribute information of the selected sampling locations and to display and analyze this information in ArcView.

An Access file and an ArcView project file were constructed to approach the goals. The Access file has at least the two basic tables, preconstructed queries and macros, and customized menu items to store the data and export the selection to ArcView. Avenue scripts were provided in the ArcView project file to provide the function of making the point coverage for the site location, and export the selection of the site location to MS Access.

The methodology of integrating the Access into ArcView is shown in the Figure 1:

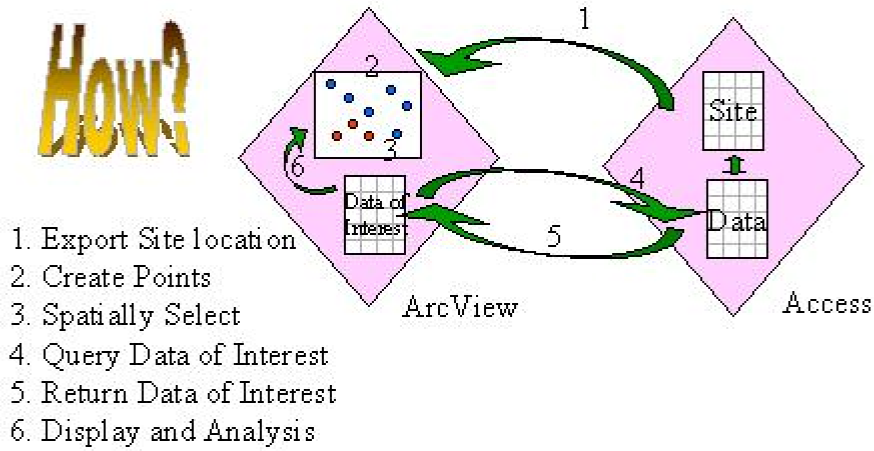


Figure 1 The Schematic View of the Integration of Access and ArcView

Computer and Data Requirement

All of the exercise can be completed using Window NT 4.0 operating system with MS Access in Office 97, and ArcView3 with Spatial Analyst extension.

The Spatial Analyst extension is used for the final analysis of the sampling data, it will not affect the intergration of Access into ArcView. So if it is not available, you can still finish most of the exercise.

The required data sets are Pantex sampling data ptxdata.xls, which is used to build the Access database, and ArcView project file AvAccess.apr, which has Avenue scripts and customized user interface to create the point coverage and export the selection.

The two files needed (ptxdata.xls and avaccess.apr) are available at CRWR Anonymous FTP site ftp.crwr.utexas.edu, under the directory pub/gisclass/accessav.

Exercise Procedure

1. Set File Directory Structure

In your working directory, create the directory AccessAv and a subdirectory ArcView under AccessAV. Copy the initial data table ptxdata.xls and ArcView Project file avaccess.apr into the AccessAv directory.

2. Build the Environmental Sampling Database

Create the Sampling Data Table

The sampling data table can be easily built by simply importing the whole original Excel file ptxdata.xls. Open Microsoft Access with a blank database, save the database in the AccessAv directory as AccessAV.mdb. A blank database with a name of AccessAV will show up. While the Tables tab is active (raised), click New to create a new table. Choose Import Table and the Import window will show up. Change the Files of type to Microft Excel (*.xls), and import ptxdata.xls. An Import Spreadsheet Wizard window will show up. Follow the wizard and let Access set up the primary key and save it as sampdata.

Open the sampdata table, you will notice for each sampling location, various types of chemicals have been sampled at various time. If we consider the sampling locations and types of chemical as the two dimensions, the time series is the third dimension, which siginificantly increases the complexity of the database, expecially for a large one.

Create the Site Locations Table

The initial data file contains both the site location files and the sampling data in Excel spread sheet format. In order for ArcView to create the point coverage of the site location, a table with only the site locations needs to be prepared. You will use query to retrieve the site locations. Click on the Querys tab in the database window, and then click on New. You will use Simple Query Wizard to create the query.

In Simple Query Wizard, choose Table: Sampdata under the Table/Query drop down list, and select only the three fields Location Code, Easting, and Northing. Your Simple Query Wizard should be like the Figure 2.

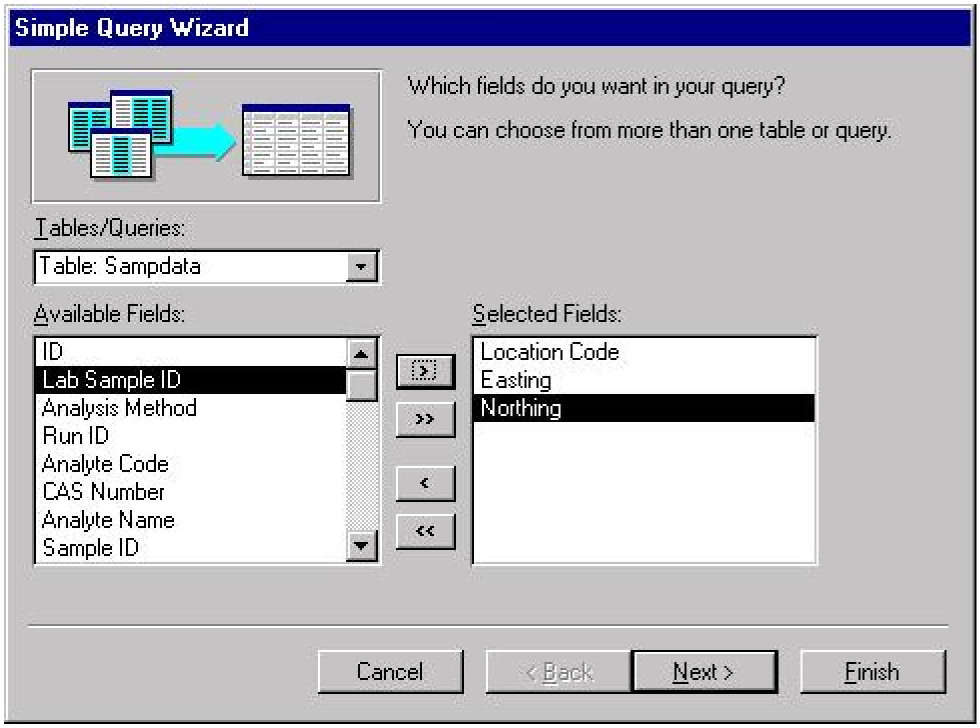


Figure 2 Simple Query Wizard for Creating a Site Table

Follow the wizard and choose the detailed query, give the name as Site Locations and finish the query building. Double click the query Site locations and open it. You'll notice that the site locations are duplicated for all of the sampling data. What we want is the single entry for each location. You can change Query Properties show only the unique value of the site location. Switch to the Design View by clicking on Design View tool, then click on the Property tool , the Query Properties window will pop up. Click on the item Unique Value and swith no to yes. The screen should look like Figure 3:

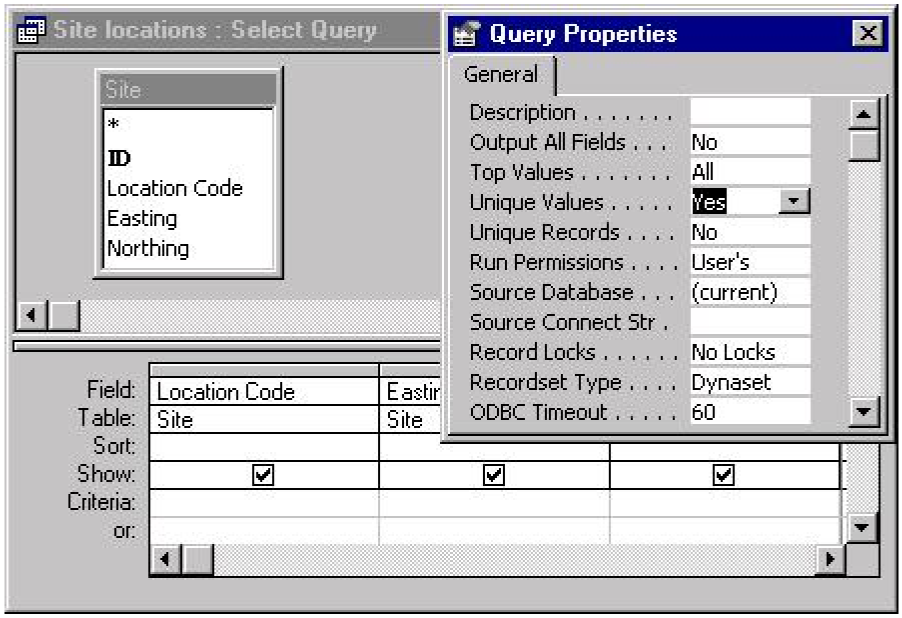


Figure 3 Setting Up the Query Properties

Then switch to Data Sheet View by clicking on the Design Sheet View tool. You will notice that the query Site Locations now only display the unique value of the sampling locations. However, there is one more thing that we want to improve- we would like to give the ID number to the locations. In order to do that, we need to export the query Site Locations as an Excel files, import as a table agian and let the Access to set up the primary key.

Choose File/Save as, change directory to ArcView subdirectory and the file type to Microft Excel 97 (*.xls), and click Export. The Excel file site_locations.xls will be saved in ArcView directory. Click on Tables Tab in Access Database window and then click on New. Import the Site locations.xls, let the Access to add the primary key and use the table name site_location. (There might be a better way to set up the primary key, which can be the future improvement for this exercise)

3. Export the Site Locations to ArcView

The table Site_Locations will be exported as a dBASE file so that it can be read by ArcView. Choose menu File/Save As, and select the option To an External File or Database. Change the directory to ArcView subdirectory, choose dBASE IV (*.dbf) in the item Save as type drop down list and give a file name Site.dbf. Then click Export, a dBASE file site.dbf will be saved in the subdirectory ArcView.

One way to speed up the process is to use a macro. We will write a macro to export the site_location table as a dBase file site.dbf to the ArcView subdirectory. Click on the Macros tab in the Database window and then click on New, a new macro design window will show up. Click on the first row under Action, and choose Transfer Database in the drop down list, a few Action Arguments will show up. Choose Export for the Transfer Type, dBase IV for Database Type, give the path of the ArcView subdirectory as the Database Name, leave the Object Type as Table, type in site_location for the Source and site.dbf for the Destination. The design window should look like the Figure 4:

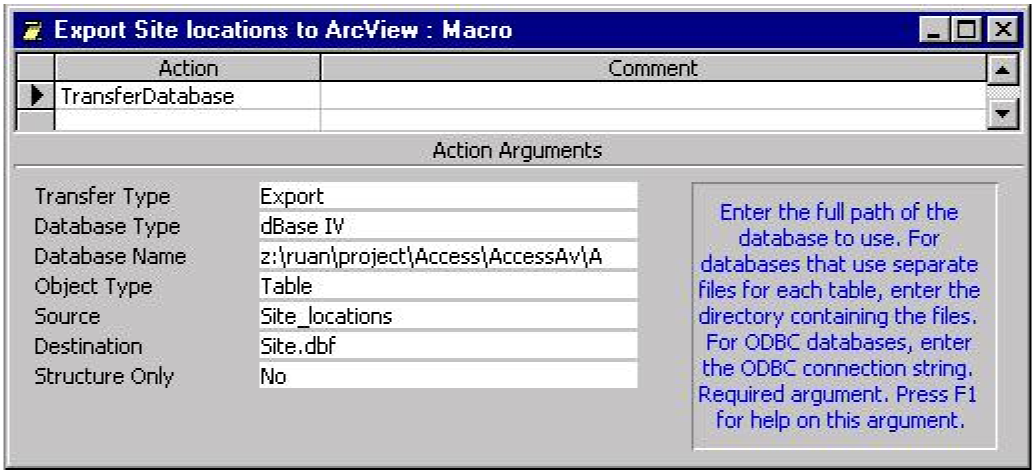


Figure 4 Design View of Macro: Export Site Location to ArcView

(Note: the DataBase Name should be your path to your ArcView directroy.)

Save the macro with the name "Export site locations to ArcView".

To run the macro, you can double-click on the name in the Database window, or use Run button. However, the easiest way to use it is to put it as a command function in the menu bar.

To customize the menu bar, choose View/Toolbars/Customize, (you can also right-click on the menu bar and choose Customize), a Customize window will pop up. Choose New Menu under Commands Categories, and drag the New Menu to the menu bar. Right-click the New Menu and rename it as AccessAv. (To delete a menu, simply drag it off the menu bar). Now click on the All Macros in the Commands Categories, and drag the macro"Export site locations to ArcView" to the item square below menu AccessAv, the marco will be added as a item of AccessAv. To run the macro, simply choose AccessAv/Export site location to ArcView.

4. Create the Point Coverage

Start the ArcView Project AvAccess.apr and save it as myAvAcc.apr in the ArcView directory.(You can keep the original AvAccess a blank copy for future uses).

Choose File/Entensions, and make sure that the Spatial Analyst extension is selected. Click once on Tables icon in the project window, and then click Add to add in the table site.dbf from ArcView directory.

An Avenue Script has been written to create the point coverage based on the latitude and longitude of the site location. This script has been implemented in a new menu item called Create Point Coverage under AvAccess.

Click once on Views icon in the Project window and double click on View1. You'll notice the menu interface of this project is different from the stardard ArcView project. An additional menu AvAccess was added. Choose AvAccess/Create Point Coverage and specify the Table with Lat/Lon is site.dbf. Choose Northing as the latitude field, Easting as the longitude field and ID as the ID field. Use the default file name sampsite and make sure the file is saved in the ArcView directory, click on OK. A shape file theme Sampsite.shp will show up in View1. Click on the side of the Sampsite.shp to make it active and click on zoom to active theme tool , the point coverage of the sampling sites will show up in

View1. Click on the Table icon , the spatial attribute information of the site as longitude and latitude will be displayed.

5. Spatially Select the Points of Interest

We can use the spatial selection tool in ArcView to select the sites of interest, and ask for a particular type of information about those sites from Access.

Click on Select Feature tool, and draw a rectangle to include the sites of interest. You can use the shift key to select any additional sites that can not be included in the rectangle. The screen should look like Figure 5:

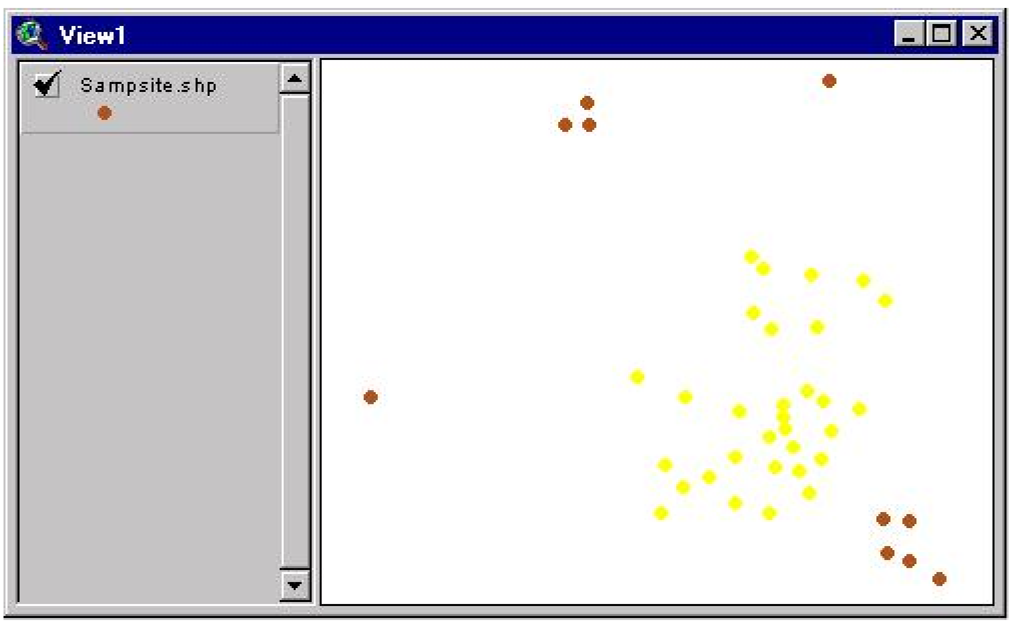


Figure 5 Spatially Select the Sites of Interest

You will export the selection as a dBase file so it can be imported into Access. Choose AvAccess/Export Selection, use the default name avselect.dbf, and save the file in the directory ArcView.

6. Import the Selection of ArcView to Access

The avselect.dbf can be imported into the Access using the same table import process as described previously, however using a simple macro can speed up the process.

Click on the Macros tab in the Database window, and then click on New, a design view of Macro1 will show up. Click on the first row under Action, and in the drop down list, choose TransferDatabase. In the Action Arguments portion, choose Import for the Transfer Type, dBase IV for Database Type, give the path of the ArcView subdirectory as the Database Name, leave the Object Type as Table, type in avselect.dbf for the Source and avselect for the Destination. Save the marco as "Import Selection from ArcView"

Choose View/Tool/Customize, and in the Customize window, click on the All Macros under Command Categories, the available macros will show up. Drag the macro "Import Selection from ArcView"as the second item to menu AccessAv.

Choose menu AccessAv/Import Selection from ArcView, you will see the table avselect show up as a new table in the database window.

7. Query the Data of Interest

As mentioned previously, the environmental sampling data are very complicated. To retrieve the useful information, you should really know what types of information that you want.

In this exercise, we will use the maximum values of the nitrate concentration for every sampling location as an example of the data of interest.

You can retrieve the data of interest by query. There are many ways to make the query work. In this exercise, we'll try to build the query from scratch.

Click on the Querys tab in the database window, and then click on New, choose Design View. Access will ask you to select the tables for the query. Hold the Ctrl key and multi-select the three tables avselect, sampdata, and site_locations, click Add and close the Show Table window. The three tables will show up in the up portion of the design view. The relationships between the tables need to be built before we can start the query. The table avselect has a same ID field with table site_locations, while the table site_locations has the same field location code as table sampdata. To connect the table avselect and site_locations, simply click on the ID field in avselect and drag it to the ID field of site_locations. You can connect the table site_locations with table sampdata by draging the field location code. You can move the table around make it easy to see the relationships between the tables.

Now double click on the following four fields: the *ID* field of the table *avselect*, the *location code* field of the table *site_locations*, and the *Analyte Code* field and *Measured value* field of the table *sampdata*. All of these fields will show up in the lower portion of the query design view. If you click on the data sheet view tool, you will see all of the sampling data for those sites specified by the table avselect will show up. Since what we want is only the nitrate concentration. So type in nitrate in the criteria row for *Analyte Code* field, and you will see in the datasheet view that only nitrate contrations show up. Now we need to specify that what we want is the maximum concentration of nitrate at each site. Switch back to design view by click design view tool , and choose menu View/Total. You will see an additional row labeled Total show up in the lower portion of the query design view. Change the Group by to Max in the Measured Value field, and check the datasheet view. You will see the query only contain the maxmium nitrate concentration for each of the sampling site. Save the query as Max nitrate concentration.

8. Export the Data of Interest to ArcView

The query Max nitrate concentration can be exported as dBASE file so that it can be added into ArcView. You can use File/Save as to export the query to the ArcView directory as dBase IV file maxnitra.dbf.

If this is a routine process, you can also create the macro to speed up the process.The process is similar to the previous macros. Be sure to add the macro to AccessAv menu for a easy use. The AccessAv menu will look like the Figure 6:

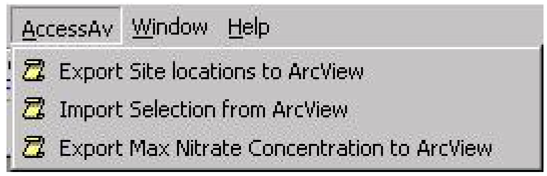


Figure 6 The customized user interface for Integration of Access into ArcView

9. Display and Analyze the Data of Interest in ArcView

The query result can be added into the ArcView project myAvAcc.apr.

Go back to the myAvAcc.apr project, hightlight the tables icon in the project window, and add in the maxnitra.dbf from the ArcView directory.

The information in the maxnitra.dbf needs to be joined with the attribute information of the theme sampsite before they can be linked to the spatial information. To join the table maxnitra.dbf to the Attribute table of Sampsite, click on the title bar of the table maxnitra.dbf and then click on the field ID, the ID field seems indented. Next, click on the title bar of the Attribute table of Sampsite and then click on the ID field. Choose Table/Join, you will see the maxnitra.dbf is joined to the Attribute table of Sampsite.

You can use the identify tool to query each of the selected sampling point and find out the maximum nitration concentration of that site. You can also label each of the site with its maxmium nitrate contration.

In this exercise, we'll try to create a contour map of the nitrate concentration of the ground water of the Pantex facility.

Choose menu Analysis/Creat Contours, and change the Output Grid Extent in the Surface Grid Specification window to Same As Sampsite.shp, click OK. Then in the Z Value Field dropdown list, choose the field *maxofmeasu*, click OK. After a few moments, you need to specify the interval, you can use the default value for the first time. ArcView will then draw the contours based on the Z Value field, in this case, it is the maximum of nitrate concentration. You will get a screen like the Figure 7:

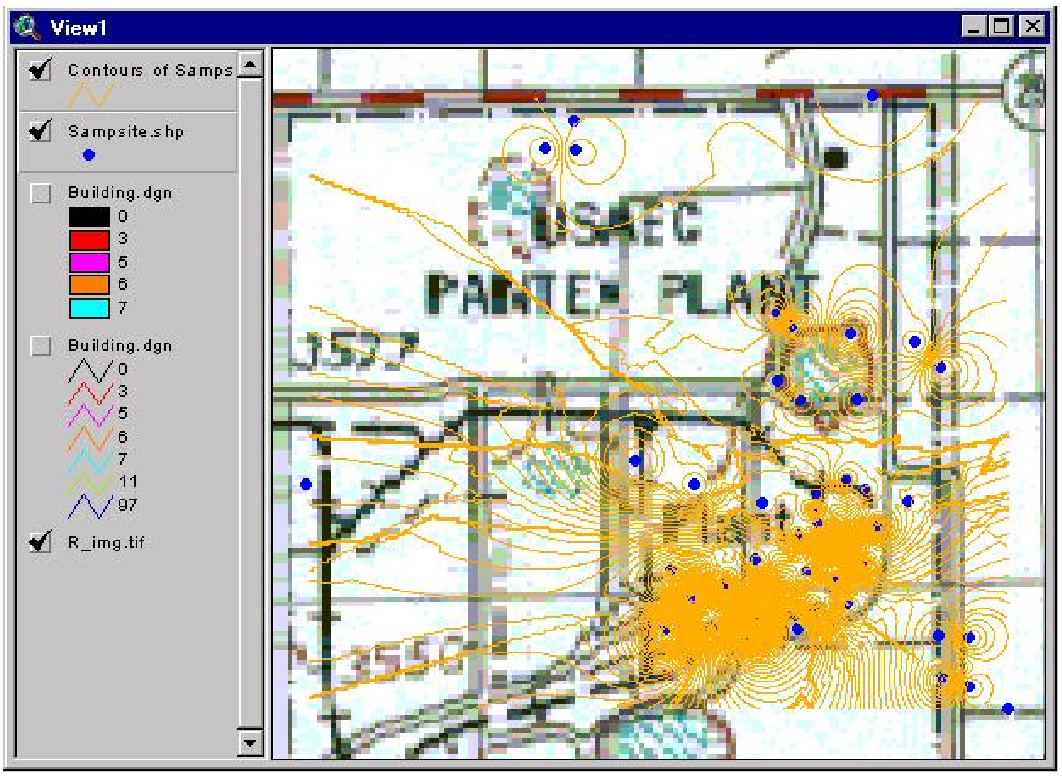


Figure 7 The Nitrate Distribution in the Groundwater of the Pantex Facility

You have successfully integrated Access with ArcView!

10. Employ the integration

You need to clean up the ArcView project before using it for a analysis of a new set sites. Delete the theme Contours of Sampsite by making it active and choose Edit/Delete Theme. Click on Tables in the project window and choose Table/Remove All Joins. Delete the table maxnitra.dbf.

Now you can select another set of points and use AvAccess/Export to export a new dBase file avselect, then in Access, delete the table avselect, which contains information of the previous selected points, and choose AccessAv/Import Selection from ArcView. Since the query max nitrate concentration has already been set up, this query now contains the maximum nitrate information of the new set of the sites (you can verify this by simply opening this query). Choose AccessAv/Export Max Nitrate Concentration to ArcView, and add in the table maxnitra.dbf to ArcView, you can display and analyze the information of this new set of sites in the same fasion.

11. Clean up

If you want to use the tables, queres and macros in the Access database AccessAv.mdb, you have to delete the avselect table before the next time you run the macro Import Selection from ArcView. The reason is that this macro will not overwrite the previous avselect table, instead it gives a new table with the name avselect1 if the previous table avselect is not deleted. Since the query maxnitra.dbf was set up to link to table avselect, the information in avselect1 will be neglected. This situation should be improved in the future.

Also be sure to delete the Contours of the Sampsite in the View1, remove all joined table and delete the table maxnitra.dbf.